Reality Sensing, Mining and Augmentation
for Mobile CitizenGovernment Dialogue
FP7-288815

D1.3 - Privacy Aware Sensor Data Storage and Miner

Dissemination level:	PU - Public
Contractual date of delivery:	Month 33, October 2014
Actual date of delivery:	Month 34, November 2014
Workpackage:	WP1 - Reality Sensing and Mining
Task:	T1.3, T1.4
Type:	Prototype
Approval Status:	PMB Final Draft
Version:	14
Number of pages:	56
Filename:	D1-3.tex

Abstract

This deliverable presents the privacy aware, server based reality sensing infrastructure for Live+Gov applications.
We review the legal and ethical aspects of privacy and perform a thorough security analysis of the Live+Gov applications. We identify six main threats to privacy and make eight recommendations for addressing these threats. These recommendations form the guidelines for several privacy enhancements to our infrastructure which we describe in this document.
Along with this document we supply the source code of the components.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

This work was supported by the EU 7th Framework Programme under grant number IST-FP7-288815 in project Live+Gov (www.liveandgov.eu)

History

Version	Date	Reason	Revised by
01	2014-07-17	Added Outline of Deliverable	Heinrich Hartmann
02	2014-09-10	Started IT Security Analysis	Heinrich Hartmann
03	2014-09-13	Added Threat Table with Conflicts and Vulnerabilities	Maximilian Meffert
03	2014-09-15	Description of IT Security Analysis	Heinrich Hartmann
04	2014-09-17	Finished Description of Actors and Interests	Maximilian Meffert
05	2014-10-08	Added chapter on Legal Aspects of Privacy	Heinrich Hartmann
06	2014-10-13	Updated 7 Types of Privacy	Heinrich Hartmann
07	2014-10-15	Updated IT Systems Description	Heinrich Hartmann
08	2014-10-21	Added Implicit Privacy Type Violations	Maximilian Meffert
09	2014-10-25	Added Description of Implementations	Heinrich Hartmann
10	2014-11-05	Extended Threat Scenarios	Heinrich Hartmann
11	2014-11-07	Added Introduction and Executive Summary	Heinrich Hartmann
12	2014-11-20	Added Spanish Privacy Law	Maite de Arana Agiretxe
13	2014-11-21	Added Summary Section and Updated Graphics	Heinrich Hartmann
14	2014-11-24	Finalization	Heinrich Hartmann

Author list

Organization	Name	Contact Information
UKob	Heinrich Hartmann	Phone: +49 261 287 2759 Fax: +49 261 287 100 2759 E-mail: hartmann@uni-koblenz.de
UKob	Maximilan Meffert	E-mail: maxmeffert@uni-koblenz.de
BIZ	Maite de Arana Agiretxe	E-mail: komunikazioa.tekniko1@biscaytik.eu

Executive Summary

This deliverable presents the privacy aware, server based reality sensing infrastructure for Live+Gov applications. Privacy protection is a crucial point for public acceptance of advanced data mining techniques in eParticipation scenarios. In order to provide such protection we perform an extensive analysis of all threats to privacy and implement a number of measures that address these threats in an adequate way.

The term ”privacy” itself is highly ambiguous and there is no common understanding what privacy really is in the literature. In Chapter 2 we discuss different legal and philosophical texts and about the subject. The main relevant legislation is the EU Data Protection Directive (cf. Section 2.1.3). The main ethical treatment is provided by Charles Fried (cf. Section 2.2.1), who defines privacy as the *control over personal information*. Furthermore, seven concrete privacy types are introduced (cf. Section 2.2.2) and the relation to collected sensor data from mobile phones is made.

On the basis of these foundations perform a thorough analysis of all risks for the privacy of a citizen. The main difficulty we are presented with is the selection of sensible measures. There are a lot of anonymization techniques in the literature and it is easy to jump ahead and implement a few of them while leaving out massive privacy gaps at other ends. Therefore a structured approach to this problem was taken.

We have chosen to use the IT Security Reference Model [10] (cf. Section 3.1), that was developed by our colleague from Prof. Grimm from University of Koblenz, as a framework to conduct a systematic analysis of all risks that are posed to the privacy of the citizens. The analysis is carried out in Section 3.2. Prof. Grimm also contributed with his expertise and guidance to the execution of this analysis.

The analysis starts by describing an abstract version of the IT Systems and the involved actors with their interests. We then describe vulnerabilities of the IT Systems and conflicts between the interests of the stakeholders. All threats that are posed to the privacy of the citizen are caused by a conflict of interest between two actors and exploit a vulnerability of the system.

We identity six main threats that include “T1. Insufficent Control Features”, “T2. Excessive Data Minig” and “T4. Surveillance” and asses their risk of occurrence. The biggest threat to the citizens privacy is posed by Insufficient Control Features of the system. As soon as collected data of Citizens is stored on the servers, all control over that data is lost if the service provider does not offer tools for the citizen to control his data.

These risks are addressed by eight recommendations, that when implemented provide an adequate protection of the citizens privacy. One surprising outcome, is that the main privacy threats can be eliminated by enhancing communication with the citizen and establishing policies and awareness for privacy aspects inside the company. Thus, a central point in the protection of the citizens privacy is a “Privacy Dashboard” that gives the citizen a maximum of control over his data. Using this dashboard he is able to view and export all data from the citizen that is currently stored in the Live+Gov system, selectively delete parts of the stored data and get information about processing applied and view the processing results themselves.

In addition, we have implemented four different anonymization methods for GPS samples, several security mechanisms for infrastructure and communication channels, moreover we wrote a concise and comprehensive privacy policy that informs the citizen about all processing steps.

The deliverable is accompanied with source code of the components in Java, Python and JavaScript.

Abbreviations and Acronyms

AIDL	Android Interface Description Language
AJAX	Asynchronous JavaScript and XML
API	Application Programming Interface
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HAR	Human Activity Recognition
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
ID	Identifier
JSON	JavaScript Object Notation
LAN	Local Area Network
REST	Representational State Transfer
SLD	Service Line Detection
SQL	Structured Query Language
UI	User Interface
URL	Uniform Resource Locator
UUID	Universal Unique Device Identifier
WP	Work Package
WIFI	Wireless Fidelity (IEEE 802.11), WLAN
WLAN	Wireless Local Area Network
XML	Extensible Markup Language

Table of Contents

List of Figures

1 Introduction

In recent days the importance of privacy protection has been amplified by the reports about the mass surveillance of ordinary citizens on a global scale by the NSA and other intelligence agencies around the world.

While aiming at the noble cause of enhancing eParticipation using mobile technologies, Live+Gov systems do process a large variety data that is potentially infringing the citizens privacy. The captured data includes personal information like name, phone numbers and email addresses and sensor data from GPS and accelerometer sensors. Also with some applications it is possible to gather images and textual input from the citizen.

While the collection of this data is necessary for providing the advanced services that Live+Gov aims to deliver, at he same time, the available raw data can be used to draw a very detailed picture of the private life of the citizen. For instance can GPS location tracking be used to reveal shopping habits (e.g. when a car seller is visited) and associations to political groups (when a meeting is attended). Accelerometer data can be used to infer medical conditions like walking disabilities. Images can contain faces of nearby persons to with whom the citizen is associated. All this data is highly sensitive to the citizens privacy and can be used against the citizen if it falls in the wrong hands.

The great importance of protecting the citizens privacy should be apparent from these examples. The European Union, as well as many other countries in the past, has set out a number of directives that regulate the collection, processing and use of privacy sensitive data. We explain the most relevant legislation in Section 2.1.

The ethical aspects of privacy have been the subject of study of many social scientists and philosophers. One scholar which is particularly relevant in our context is Charles Fried. He investiages, why we are intuitively so sensitive to violations of our privacy. For him privacy is not asserted as an intrinsic value by itself, he rather sated:

> Privacy is not simply an absence of information about us in the minds of others; rather it is the control we have over information about ourselves.

Fried‘s study on the understanding of privacy provided a great contribution to the research on the same term in philosophy and computer science and despite the fact his text was published in 1970, he already included technologies to its viewpoint (like location monitoring) that are particularly relevant to our Context. We explain his theory in 2.2.1 and follow his definition of privacy in this document.

In this document we perform a thorough analysis of how to protect the privacy of a citizen and present several implementations that form the core of our privacy aware Sensor Data Storage and Mining Infrastructure.

We identify six main threats for the citizens privacy, and derive eight recommendations that should be followed in order to reduce the associated risks for hazards. This analysis form the guideline for the selection of 9 measures that were implemented in our system and documented in Chapter 4.

2 Legal and Ethical aspects of Privacy

As privacy is a very general and hard to grasp term, we need to fix a definition of privacy that is suitable for our needs. As background information we include an overview about historical treatments of privacy as well as legal regulation of privacy in the European Union. Based on this we propose a definition of privacy as *control over personal data*, and introduce seven privacy types that give specify the term *personal data* in the context of mobile sensor data collection.

2.1 Legal Aspects of Privacy

In the following sections we outline the most relevant legislature regarding privacy protection, from on European angle. Of particular importance is the EU Directive 95/46/EC that is discussed in section 2.1.3.

2.1.1 European Convention on Human Rights

The *Europen Convetion on Human Rights* [15] was created by the members of the *Council of Europe* (CoE) in 1953 as part of the aftermath of the second world war. It formulates universal rights of citizens against the state authority. All member states of the Council of Europe, which includes all EU members, have incorporated this convention into their national law.

Article 8 of the ECHR contains a protection of personal data.

> (1) Everyone has the right to respect for his private and family life, his home and his correspondence.
>
> (2) There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

The protection of personal data under Article 8 is not an absolute law but must be considered in relation to other laws such as the freedom of expression (ECHR, Art. 10). The *European Court of Human Rights* (ECtHR) overseas the implementation of this directive and has adopted a rather broad interpretation of the article.[1]

In particular State actions of searching a persons home, gathering and storing information in a secret police file, and stopping a prisoner's communication have been ruled to interfere with Article 8.

The Live+Gov systems can violate this fundamental human right if it is used to gather information about the private life without the knowledge and consent of the citizen.

[1] http://en.wikipedia.org/wiki/Article_8_of_the_European_Convention_on_Human_Rights

2.1.2 OECD Guidelines and CoE Convention 108

The emergence of information technologies that allow automated processing of personal data made more detailed rules for safeguarding the privacy of citizens necessary. In the light of this developments the OECD [14] issued Guidelines on Protection of Privacy in 1980 which establish the following basic principles of data protection:

1. Collection Limitation Principle. There should be limits to the collection of personal data and any such data should be obtained [...] with the knowledge or consent of the data subject.
2. Data Quality Principle. Personal data should be relevant to the purposes for which they are to be used, and, [...] accurate, complete and kept up-to-date.
3. Purpose Specification Principle. The purposes for which personal data are collected should be specified [...] and the subsequent use limited to the fulfillment of those purposes [...].
4. Use Limitation Principle. Personal data should not be disclosed [...] except: a) with the consent of the data subject; or b) by the authority of law.
5. Security Safeguards Principle. Personal data should be protected by reasonable security safeguards [...].
6. Openness Principle. [...] Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.
7. Individual Participation Principle. An individual should have the right
   - to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him;
   - to have communicated to him, data relating to him [...]
   - [...] to have the data erased, rectified, completed or amended.
8. Accountability Principle. A data controller should be accountable for complying with measures which give effect to the principles stated above.

Where the following definitions are understood.

- *personal data* means any information relating to an identified or identifiable person (*data subject*). An identifiable person is one who can be identified either directly, by reference to a name or identification number or indirectly, by one or more factors which make it possible to find out who the data subject is by conducting further research.
- a *data controller* means any party who is competent to decide about the contents and use of personal data.

The OECD guidelines were adopted by EU law in the Strassbourg Convention 108 of the Council of Europe [16] in 1981.

Relevant for our investigations is in particular, that sensor data collected from mobile devices is considered personal data if the individual can be identified based on the data by any direct or indirect means. Therefore the above principles apply.

2.1.3 EU Data Protection Directive

The *Directive 95/46/EC* of 1995 [20] is the fundamental regulation of data protection in the European Union. It was designed to give further substance to the Convention 108. In particular it makes the creation of an independent supervisory authority necessary (Art. 28). As this text is the main legal basis for our investigation we review it in detail. In Section 2.1.4 the status of the implementation is discussed in some example cases.

The interpretation and oversight of the Directive lies at the Court of Justice of the European Union (CJEU). A summary of the most important rulings can be found in Handbook on European data protection law [8].

The ambition of the Directive is stated at the very beginning.

> (Art. 1.1) In Accordance with this Directive the Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data.

The Directive follows the definition of personal data, data subject, and data controller of the OECD Guidelines (cf. Art. 2). Furthermore it introduces the role of a *processor* as a person which processes, by wholly or partly automatic means (Art 3.), personal data on behalf of the controller.

Data collection of intelligence agencies and the police is not restricted by the Directive (Art. 3).

The most important articles are summarized in the following paragraphs. In order to keep the text more readable we have simplified the formulation by removing the formulation "Member States shall provide that ..." from each paragraph.

- (Art. 6.1.a) Personal data must be processed fairly and lawfully.

  [The meaning of the term 'fair processing' is entailed in the following articles. In particular transparency of processing, information of the data subject, and right to access the data are implied (cf. [8])]

- (Art. 6.1.b) *Specification of Purpose.* Personal data must be collected for specified, explicit and legitimate purposes and not be processed in a way incompatible with this purposes.

  Further processing of data for historical, statistical or scientific purposes shall not be considered incompatible [...].

- (Art. 6.1.c) Personal data must be adequate, relevant and non-excessive in relation to the purposes [...].

- (Art. 6.1.d) *Accuracy.* Personal data must be accurate and [...] every reasonable step must be taken to ensure that data which is inaccurate or incomplete [...] are erased or rectified.

- (Art. 6.1e) Personal data must be kept in a form which permits identification for no longer than is necessary [...].

In addition to strengthening the OECD Data Quality Principle in Article 6, the Directive requires a strong form of consent of the data subject, before any collection or processing of data can take place.

- (Art. 7). Personal data may be processed only if:

    (a) the data subject has unambiguously given his consent; or

    (..) [exceptions are made for performance of a contracts with the data subject, legal obligations of the controller, to protect vital interests of the data subject, the public interest and]

    (f) the processing is necessary for purposes of legitimate interests pursued by the controller or by a third party to whom the data are disclosed, except where such interests are overridden by the interests [...] of the data subject which require protection under Art. 1(1).

- (Art. 8.1) *Special Categories.* Member Stats shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life. [Art. 8.2 lists several exceptions, including explicitly explicit consent of the data subject.]

The OECD Participation Principle is addressed and extended in the following articles.

- (Art. 10, 11) The controller [...] must provide a data subject [...] with the following information:

    (a) the identity of the controller [...]

    (b) the purpose of processing for which the data are intended;

    (c) any further information [...] that is necessary to guarantee fair processing in respect of the data subject. [This includes in particular the recipients of the data, the existence of the right to of access and the right to rectify the data.]

- (Art. 12) *Right to Access.* Every data subject has the right to obtain from the controller
    - (a) [...] confirmation as to whether or not data relating to him are being processed [...], communication to him [...] the data undergoing processing and any available information as to their source, knowledge of the logic involved in any automatic processing [...];
    - (b) the rectification, erasure or blocking of data the procession of which does not comply with the provisions of this Directive [...];
    - (c) notification to third parties to whom the data have been disclosed [...].

A similar form of the Security Safeguard Principle is contained in Article 17.

- (Art. 17) *Security of processing.* The controller must implement appropriate technical and organizational measures to protect personal data against accidental loss, alteration, unauthorized disclosure or access [...] and against all other forms of unlawful processing.

  [The level of security has to be balanced against the risk of processing.]

Furthermore the Directive demands all data processing to be reported to a supervisory authority. This obligation can be lifted if an internal data protection official is appointed, who is responsible in particular for keeping a processing register, that has to be made available on request.

- Notification (Art. 18. 1). The controller [...] must notify the supervisory authority referred to in Article 28 before carrying out any wholly or partly automatic processing operation [...].

- (Art. 18.2). Simplification or exemption from notification may be provided only under the following conditions: [...] where the controller [...] appoints a personal data protection official, responsible in particular:
    - for ensuring in an independent manner the internal application of the national provisions taken pursuant to this directive.
    - for keeping the register of processing operations carried out by the controller, containing the items of information referred to in Article 21 (2)

  thereby ensuring that the rights and freedoms of the data subjects are unlikely to be adversely affected by the processing operations.

- (Art. 19.1) The information to be given in the notification shall include at least:
    (a) the name and address of the controller [...];
    (b) the purpose [...] of processing;
    (c) a description of the category [...] of the data [...];
    (d) the recipients [...] to whom the data might be disclosed;
    (e) proposed transfer of data to third countries;
    (f) a general description allowing a preliminary assessment to be made of the appropriateness of the measures taken pursuant of Article 17 to ensure the security of processing;

- (Art. 21.2) *Publication.* A register of processing information [...] shall be kept at the supervisory authority. The register shall contain at least the information listed in Article 19.1 (a) to (e).

  The register may be inspected by any person.

- (Art. 21.3). In relation to processing operations not subject to notification, that controllers [...] make available at least the information referred to in Article 19.1 (a) to (e) in an appropriate form to any person on request.

- (Art. 28) *Supervisory Authority.* Each Member State shall provide that one or more public authorities are responsible for monitoring the application [...] of this Directive.

  These authorities shall act in complete independence in exercising this functions entrusted to them.

Data transfer to third countries outside of the EU requires those countries to have an adequate level of data protection.

- (Art. 25.1) *Transfer to third countries.* The transfer to a third country of personal data [...] may take place only if [...] the third country in question ensures an adequate level of protection.

The European Commission can decide whether a third country ensures an adequate level of protection (Art 25.6). The USA, is not considered to do so. Data exchange between the US and EU is possible under the Safe Harbor regulation [5]. Further Exceptions are provided by the Passenger Name Record Agreement [18].

2.1.4 Implementation of the Data Protection Directive

Germany. Germany implements Directive 95/46 with the *Bundesdatenschutzgesetz (BDSG)* of 2001. However, Germany has violated the directive in two points:

1. The BDSG has become effective three years too late, thus the EC filed a treaty violation proceeding against Germany.
2. The BDSG does not implement independent supervisory authorities. The Bundesdatenschutzbeauftragter is subordinate to the Ministry of Interior. Although he is not subject to technical oversight (*Fachaufsicht*), he is subject to staff supervision by the government (Rechtsaufsicht, Dienstaufsicht) and budget oversight by the ministry. In March 2010 Germany was found guilty of violation of Directive 95/46 by the ECJ.

The states of Germany have their own implementation of Directive 95/46 (*Landesdatenschutzgesetze*). Federal public authorities are only bound to their federal law. Churches are not subject the BDSG.

United Kingdom. The UK implements Directive 95/46 with the *Data Protection Act 1998* (DPA).

The act is known for its high complexity: a manual record of phone numbers for business purposes could be hold subject to the DPA. Although the act seems to fully cover the directive. Even higher restriction apply for *"sensitive personal data"* (race, ethnicity, politics, religion, trade union status, health, sex life or criminal record), i.e. consent must be given freely and has to be explicit.

The Freedom of Information Act 2000 modified the act for public bodies and authorities, and the Durant case modified the interpretation of the act by providing case law and precedent.

Spain. Spanish Law 15/1999, Ley Orgnica de Poteccin de Datos de Carcter Personal (LOPD), is the implementation of the Directive 95/46. Although the compliance of some of the articles from the Spanish law with the Directive has been studied by the ECJ, the LOPD implements this directive in all aspects. In the case of Spain there are several different supervisory authorities, the Spanish Agency for Data Protection (AEPD, Agencia Espaola de Proteccin de Datos) and the Agencies of the Basque Country and Catalonia (Agencia Vasca de Proteccin de Datos - AVPD, Autoridad Catalana de Proteccin de Datos-ACPD). In the case of the AVPD, this agency is particularly committed to public authorities, and does not register corporate personal data files of any type.

United States of America The USA do not implement the directive, nor is there any obligation for them to do so. However, companies subject to US jurisdiction can be certified to comply with the seven principles enforced by Directive 95/46. Thus, those companies will act as *safe harbors*. Without certification foreign companies are not allowed to store and process customer data in their country.

2.1.5 EU Charter of Fundamental Rights

The Treaty of Lisbon of 2007, which introduced the fundamental functioning of the European Union includes the *Charter of Fundamental Rights of the Europen Union* [19]. This Charter summarizes the full range of civil, political and economic rights of EU citizens and contains the following two articles that safeguard the privacy of the citizen.

> Article 7. Respect for private and family life.
> Everyone has the right to respect for his or her private and family life, home and communications.
>
> Article 8. Protection of personal data.
>
> (1) Everyone has the right to the protection of personal data concerning him or her.
>
> (2) Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
>
> (3) Compliance with these rules shall be subject to control by an independent authority.

Although the Charter does not extend the pre-exiting Directive 95/46/EC, it amplifies the importance of this privacy protection as a fundamental right.

2.2 Definition and Types of Privacy

In this section we introduce our definition of privacy which as *control over personal information* following C. Fried. As this description is still on a very high level we go on to specify seven different privacy types in the next section. These privacy types are then analyzed for their implications in the presence of different kinds of sensor data.

2.2.1 Defining Privacy

Defining privacy is a challenge which seems impossible. This is well put to words by Serge Gutwirth, who notes:

> The notion of privacy remains out of the grasp of every academic chasing it. Even when it is cornered by such additional modifiers as ‘our’ privacy, it still finds a way to remain elusive. [12]

Many researchers seem to only “focus on the ways in which privacy can be infringed” [7]. Thus they invest a great amount of work in defining threats to prohibit instead of describing why privacy is so valuable to us. This way around one could derive measures to ensure and secure its value [7]. One particular scholar who does this in the context of digital monitoring is Charles Fried. He questioned, why we are intuitively sensitive to violations of privacy. But he did not assert privacy as an intrinsic value by itself, he rather sated:

> Privacy is not simply an absence of information about us in the minds of others; rather it is **the control we have over information about ourselves**. [9]

According to Fried privacy is a *rational context*. Which means that one is aware of such context’s existence during a rational action. If we are to share private information with others, we are most likely aware that this action is privacy related. In that case we are able to selectively disclose information along the two dimensions of quantity and quality. This leads Fried to his thesis, that privacy is one's ability to create and modulate his or hers social relationships, namely: friendship, love and trust. [9]

Depending on the conversation partner, we change the degree of intimate information we share if it is a total stranger, colleague, close friend or lover. With close friends or lovers we share information of great intimacy we do not share with anyone else. Moreover, we trust those persons to not reveal information about us to others by respecting their privacy. Trust needs the possibility of unknown failure. If we would constantly monitor our partners, they cannot fail unnoticed nor can they willingly share that information with us. Thus they could not trust us anymore. [9]

So privacy or the its possibility, according to Fried, is the foundation of our core relations: friendship, love and trust. And thus it is valuable, because those relations are essential to human society [9]. Using his anatomy of privacy as foundation of our analysis is suitable because of two points. At first, Fried's study on the understanding of privacy provided a great contribution to the research on the same term in philosophy [6] and computer science [2]. Secondly, despite the fact his text was published in 1970, he already included technologies to its viewpoint that did not only monitor location, but also record biometric data.

- Subject - Trusted Person - Third Party which shall NOT obtain sensitive information.

- Control over information in law

2.2.2 The Seven Types of Privacy

In Fried's definition of privacy as control over "information about onself", the specification of what constitues such information remains open. There is a vast amount of information that relates to a person and we need to get a better understanding in order to perform a thorough analysis. To this end we use the the categorization by Friedewald, Finn and Wright [7] called the *Seven Types of Privacy*. The seven types of privacy are an extension to the four types of privacy by Roger Clarke [4], which are:

- Privacy of the Person
- Privacy of Personal Behaviour
- Privacy of Personal Communication
- Privacy of Personal Data

It is important to note, that this categories do not form no taxonomy, since the categories are not mutually exclusive. For instance a written email is considered personal communication as well as personal data stored on a computer.

Moreover, Friedewald et al. argue that Clarke's taxonomy is outdated and no longer adequate in order to describe the privacy aspect of our modern, technology driven, world. In order to address this shortcoming they extend the former four to the now introduced seven types privacy as follows:

1. **Privacy of the Person** This privacy type is generally concerned with one could best understand as biometric privacy. Friedewald et al. paraphrase it as *"[...] the right to keep*

body functions and body characteristics [...] private". This includes but is not limited to measures like weight, height or shoulder width; biometric identifiers like fingerprints and DNA sequences; or medical conditions such as limping or having a cold.

2. **Privacy of Behaviour and Action** This privacy type is concerned with one's activities in public as well as in private spaces. It includes but is not limited to religious practices, political activities and sexual preferences or habits.

3. **Privacy of Communication** This privacy type is concerned with one's communication in a broad sense. It includes written correspondence, but also conversations conducted either vis-a-vis or via electronic devices. Friedewald et al. put it as the right to free discussion without unknown interception by third parties.

4. **Privacy of Data and Image** This privacy type is concerned with the secrecy of personal data, especially its automatic disclosure to other individuals and organizations. It includes data such as paychecks, insurance information or records of public administration. However, it also refers to pictures taken without consent and digital identifiers like IP addresses or social security numbers.

5. **Privacy of Thoughts and Feelings** This privacy type is the counterpart to Privacy of the Person like body and mind are counterparts of one another. Comparable to the Privacy of Data and Image, Friedewald et al. state that one's thoughts and feelings must not be automatically revealed to others. This could simply happen by the disclosure of one's diary or by technologies which allow emotion detection through biometric means. One's body temperature or iris reflexes might infer stress or excitation.

6. **Privacy of Location and Space** This privacy type is concerned with one's movements in public spaces and the protection of private spaces. Friedewald et al. qualify the first dimension as one's right to move without being identified, tracked or monitored. The second dimension is qualified as one's general right to solitude, especially the right to the inviolability of the home.

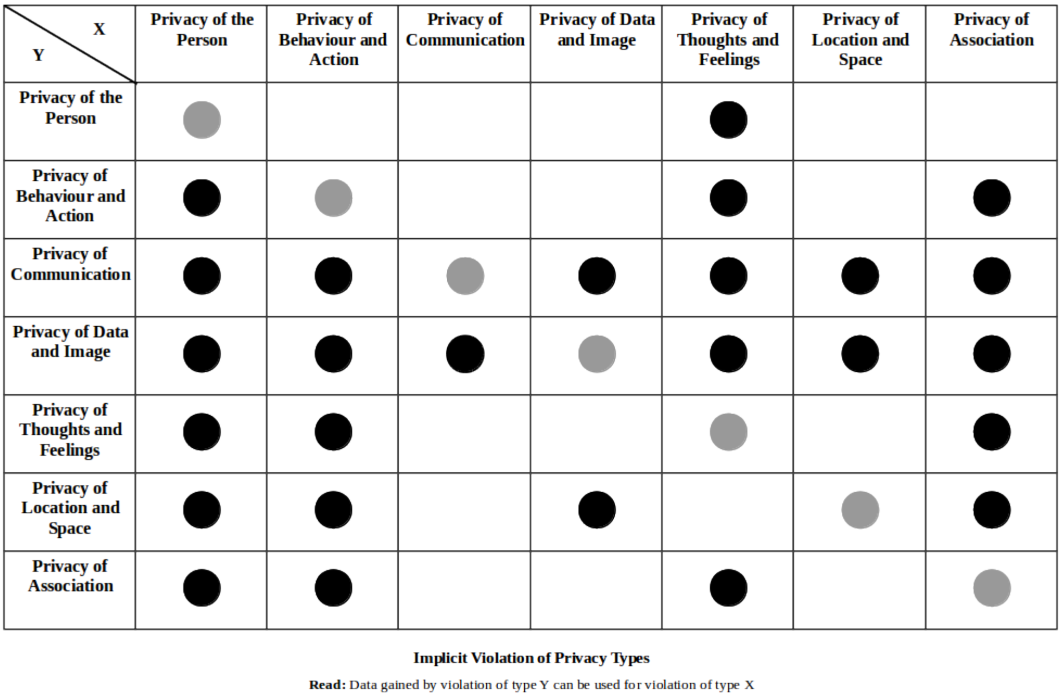
7. **Privacy of Association** This privacy type is also put as group privacy. Friedewald et al. state that one must have the possibility with whomever without being recorded. Associations like friends or organizations such as political parties must not automatically be recorded because one associates with them, and vice-versa.

2.2.3 Privacy Type Inferences

The above types of privacy types are not disjoint to each other. For example if the location of a citizen is known it is possible to infer information about political associations (e.g. his visits to a party meeting). In this section we explore possibilities how personal data can be inferred from each other. Our findings are summarized in Figure 1.

1. Privacy of The Person

The Privacy of The Person is concerned with one's biometric privacy. If this type is violated, following implicit violations are possible:



X / Y	Privacy of the Person	Privacy of Behaviour and Action	Privacy of Communication	Privacy of Data and Image	Privacy of Thoughts and Feelings	Privacy of Location and Space	Privacy of Association
Privacy of the Person	○				●		
Privacy of Behaviour and Action	●	○			●		●
Privacy of Communication	●	●	○	●	●	●	●
Privacy of Data and Image	●	●	●	○	●	●	●
Privacy of Thoughts and Feelings	●	●			○		●
Privacy of Location and Space	●	●		●		○	●
Privacy of Association	●	●			●		○

Implicit Violation of Privacy Types

Read: Data gained by violation of type Y can be used for violation of type X

Legend: The matrix above shows the following relation: *Type X can be implicitly violated by the violation of type Y.*

- The X-Axis shows the Seven Types of Privacy according to Friedewald et al., which could be violated implicitly
- The Y-Axis shows the Seven Types of Privacy according to Friedewald et al., which have been violated explicitly
- Big black bullet points denote, that an implicit violation is possible
- Big grey bullet points only denote, that the relation is reflexive (*a R a*). They are only shown for completeness sake and not discussed further, because they denote a trivial fact.

Figure 1: Implicit Privacy Violation Matrix

(1-5) Privacy of Thoughts and Feelings. Some psychological diseases (e.g. depression) have physiological impact. Such physiological patterns could be detected.

2. Privacy of Behaviour and Action

The Privacy of Behaviour and Action is concerned with one's social, political, religious, sexual, etc. activities. If his type is violated, following implicit violations are possible:

(2-1) Privacy of The Person. Religious practices which include body modifications (e.g. circumcision).

(2-5) Privacy of Thoughts and Feelings. Social activities in general depend on a certain intellectual attitude. Such an activity is the expressions of such an attitude.

(2-7) Privacy of Association. Recording religious, political or sexual activities can reveal association with churches, political parties or sexual partners.

3. Privacy of Communication

The Privacy of Communication is concerned with not having such communication (correspondence or vis-a-vis) intercepted. This is very broad type of privacy. Depending on the contents of the intercepted communication every other type can be violated:

(3-1) Privacy of The Person. Communication about body characteristics.

(3-2) Privacy of Behaviour and Action. Communication about social activities.

(3-4) Privacy of Data and Image. Communication containing one's passwords or other sensitive data.

(3-5) Privacy of Thoughts and Feelings. Communication of thoughts and feelings, e.g. wiretapping a flirt or a catholic confession ritual.

(3-6) Privacy of Location and Space. Interception of face-to-face communication is only possible if one's location and space is violated (wiretapping).

(3-7) Privacy of Association. Communication about one's associations (family members, churches, etc.).

4. Privacy of Data and Image

The Privacy of Data and Image is concerned with one's data not being automatically available to others. This also is a very broad type of privacy. Depending of the data or image contents every other type can be violated:

(4-1) Privacy of The Person. Images or stored biometric information reveal one's physical characteristics.

(4-2) Privacy of Behaviour and Action. Images or diaries can reveal one's social activities.

(4-3) Privacy of Communication. Modern communication systems usually contain some sort of archive function, e.g. E-mail clients do not automatically delete messages. Such messages are data and reveal one's communication.

(4-4) Privacy of Thoughts and Feelings. Images can show one's emotional state.

(4-5) Privacy of Location and Space. Images can reveal one's location, e.g. making a picture in front of the Eifel Tower.

(4-6) Privacy of Association. E-mail data can also reveal association.

5. Privacy of Thoughts and Feelings

The Privacy of Thoughts and Feelings is concerned with keeping such thoughts and feelings secret. If this type is violated, following implicit violations are possible:

(5-1) Privacy of The Person. Thoughts and feelings can reveal medical conditions.

(5-2) Privacy of Behaviour and Action. Thoughts and feelings can reveal a certain attitudes which create a foundation for certain social activities.

(5-7) Privacy of Association. Thoughts and feelings can reveal individual association, e.g amorous feelings for a certain person.

6. Privacy of Location and Space

The Privacy of Location and Space is concerned with one's right to move freely without being tracked and one's right to private places. If this type is violated, following implicit violations are possible:

(6-1) Privacy of The Person. Frequently visited doctors can reveal certain medical conditions, if such doctors are known specialsts. In general it could imply illness.

(6-2) Privacy of Behaviour and Action. Frequently visited places in general can reveal association and hence implies social activities.

(6-3) Privacy of Communication. If one's location is knwon, it is possible to intercept (wiretap) one's communication. This also may violate the right to private spaces.

(6-4) Privacy of Data and Image. If one's location is known, it is possible shoot pictures.

(6-5) Privacy of Thoughts and Feelings. Frequently visited persons may imply certain thoughts and feelings, e.g. having a mistress.

(6-6) Privacy of Association. Frequently visited places can reveal associations simply by searching in maps or yellow-pages.

7. Privacy of Association

The Privacy of Association is concerned with one's right to associate with whomever one wants, without that association having recorded. If this type is violated, following implicit violations are possible:

(7-1) Privacy of The Person. Association with tattoo artists could imply having tattoos or other body modifications.

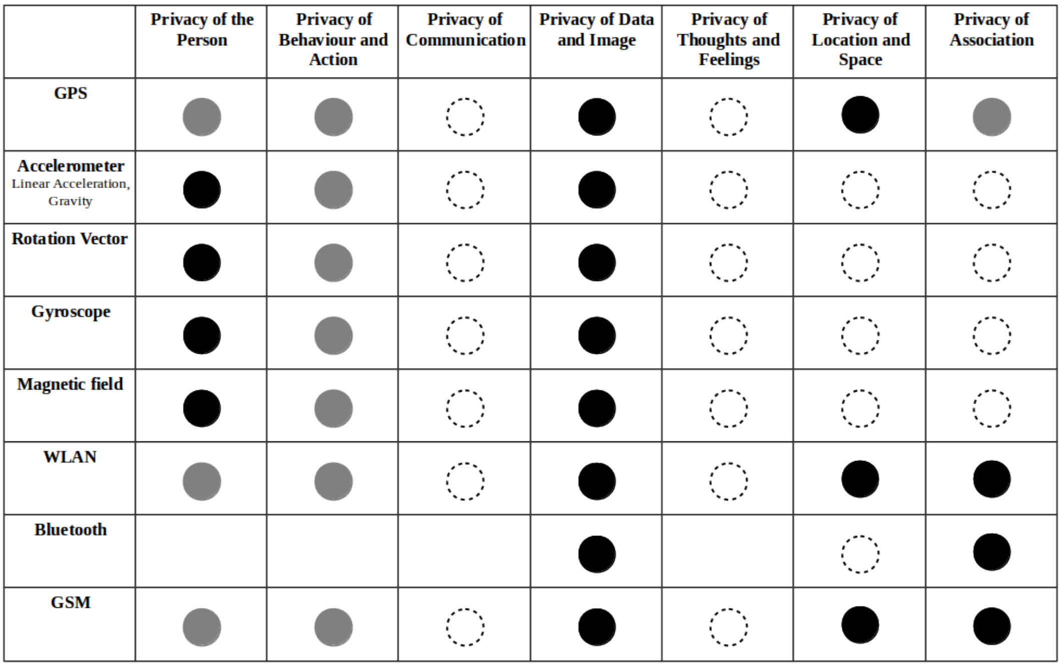
(7-2) Privacy of Behaviour and Action. Association with churches or political organizations could imply certain activities.

(7-5) Privacy of Thoughts and Feelings. Association with churches of political organizations could imply a certain intellectual attitude.

2.2.4 Sensor Data Privacy Impact

In this section we analyze the impact of the disclosure sensor data and certain processing results to the citizens privacy. This analiyisis builds upon the preceeding analyis, but is more focused on the concrete type of data vailable. For example, the disclosure of Service Line Detection results, does violate the Privacy of Location and Space but not the Privacy of Association that is violated by general GPS tracking.

GPS Data. The GPS sensor gives the current longitude and latitude, the current global position of the mobile device and its carrier, although there is some artificial inaccuracy within civil



	Privacy of the Person	Privacy of Behaviour and Action	Privacy of Communication	Privacy of Data and Image	Privacy of Thoughts and Feelings	Privacy of Location and Space	Privacy of Association
GPS	◐	◐	◌	●	◌	●	◐
Accelerometer Linear Acceleration, Gravity	●	◐	◌	●	◌	◌	◌
Rotation Vector	●	◐	◌	●	◌	◌	◌
Gyroscope	●	◐	◌	●	◌	◌	◌
Magnetic field	●	◐	◌	●	◌	◌	◌
WLAN	◐	◐	◌	●	◌	●	●
Bluetooth				●		◌	●
GSM	◐	◐	◌	●	◌	●	●

Figure 2: Live+Gov Implicit Sensor-Privacy Matrix

use. Threefore, the collection of GPS data violates directly the citizens privacy of Location and Space, and the privacy of Data an Image.

By inference we also get implicit violations of the Privacy of the Person, Privacy of Behavior and Action and Privacy of Association.

Motion Sensors. Accelerometer, Rotation Vector, Gyroscope and Magnetic field sensor measure the physical movement of the mobile device on all three axes. If the mobile device is carried "normally" its safe to say that those sensors also measure the moments of its carrier. So his privacy is infringed regarding biometric behaviour, as it is captured automatically Privacy of Data and Image is trivially threatened because here sensor data is individual data, a priori.

By inference we also get implicit violations of the Privacy of Behavior and Action.

Network Sensors. The GSM and WLAN sensors reveal the position of the mobile device and its carrier, when used in connection with external databases. The GSM sensor gives the exact cell, the mobile device has registered with at the current moment.

The Bluetooth sensors record lists of the bluetooth clients in the direct neighbourhood. Since those clients are usually moving, inference of the position is usually not possible. Instead, bluetooth clients carried by a third person may infringe the Privacy of Association.

A similar argumentat applies to WLAN snsors. If one frequently connects with an organizational wireless network, e.g. an university network, an association can be deduced (student or staff).

Human Activity Recognition. The detection and collection of human activities like walking, standing and running, can interfer with the Privacy of the Person, e.g. since these movement patterns can be indicators for a person's health. Also, trivially, Privacy of Data and Image is violated.

Service Line Detection. The detection and collection of the service line the user currently uses in the public transportation system allows inference of the location of the user, at least when entering and leaving those service lines at e.g. bus stops. Implicit privacy violations apply accordingly.

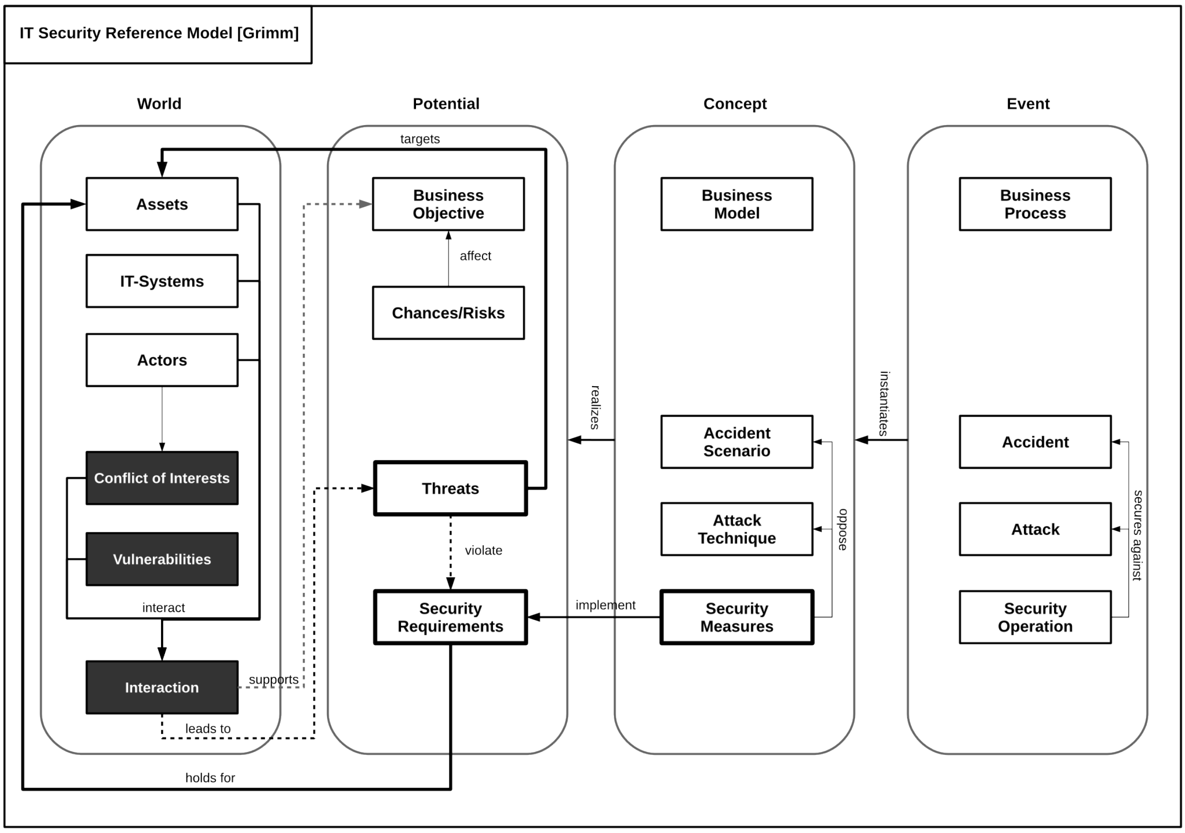


Figure 3: The IT Security Reference Model (Grimm)

3 Privacy Analysis of Live+Gov Systems

The goal of this chapter to analyze and identify the threads to personal privacy that are posed by collecting, storing and processing sensor data from mobile phones. We derive concrete privacy protection measures that address the main risks involved with handling such data.

The complexity of our systems and the variety of threads make a great number of counter measures plausible. We approach this complexity with the aid of a general security analysis model developed in [10]. We give a brief introduction to this model and perform a IT Security Analysis with respect to the privacy asset for our system.

3.1 IT Security Analysis according to Grimm et. al

We follow the Reference Model for IT Security Analysis as described in [10]. It supersedes earlier efforts by e.g. [1]. The reference model consists of a *model* and a *procedure*. The model aims to organize common security terminology in a reasonable and practical way. The procedure describes a method to analysis the IT system based on that model. In this section we give a brief overview over the reference model.

3.1.1 Model

The model is depicted in Figure 3. It is organized in four views (round boxes) that contain a number of components (rectangular boxes).

The *world view* contains all components describing the current state. It consists of the following components:

- **Actors.** All identifiable stakeholders of the system under study. Typical actors include, users, developers, clients and externals.
- **Assets.** Things of value to one or more stakeholders. The value can be hard (money, data, etc.) or soft (trust, privacy,etc.). In our case the only asset we are concerned with is the privacy of the citizen.
- **IT-Systems.** The relevant IT-Systems under study. This encompasses hardware (e.g. servers, network infrastructure), as well as software and third party services. The level of granularity has to be detailed enough to express all possible threats to the assets at stake.
- **Conflicts of Interests.** Different actors have different interests which can be in conflict which each other. These conflicts of interest are the origin of all attacks to the system.

  A typical conflict is the *Criminal-User-Conflict*: A user wants to keep control over their private data. A criminal wants to gain money. The possibility of selling private data (user profiles) to advertisers, renders both interests conflicting.
- **Vulnerabilities.** All identifiable weaknesses in the IT-System.

  In the example of the criminal-user-conflict, the criminal has to exploit a vulnerability, e.g. a weak password, to gain access to the private data about the user.
- **Interactions.** This point captures all possible interactions between assets, IT-Systems, humans and vulnerabilites. It is described in more detail in the next view.

The *potential view* displays the intended and unintended interactions of the components in the world view. The intended interactions support the underlying business objectives. Unintended interactions lead to threats. The potential view consists of the following components.

- **Business Objectives.** Interaction of IT-system and actors that realize a business goals of the system owner.
- **Threats.** A threat is a potential interaction that destroys or harms assets of the system. Concrete realizations of threats can be *attacks* or *accidents*. Attacks are executed by an actor in response to a conflict of interest. Accidents are harmful interactions that are not willfully caused by an actor.
- **Chances/Risk.** Evaluation of chances and risks associated to the business objectives and threats. The risk associated to a threat is its expected loss. A chance associated to a intended interaction is its expected gain.

  In the case that, the loss can be quantified monetary, and the likelihood of occurrence of a threat can be modeled probabilistically, the risk is given by the product

  $$\text{risk} = \text{loss} \cdot P[\text{threat}].$$

In practice such a quantitative risk evaluation is often not possible, and a qualitative, heuristic, analysis is performed instead.

- **Security Requirements.** A set of interactions (e.g. threats) that shall not occur within the system in order to achieve its business objectives. Security requirements are targeted to protect one or more assets.

  An example of a security requirement is that a given communication channel shall not be infringed by externals.

The *concept view* is a realization of the potential view of the system. It specifies important interactions that require further planning. It contains the following components:

- **Business Model.** The plan to achieve business objectives.
- **Accident Scenario.** A concrete outline of an interaction that leads to an accident. In particular the asset under threatened and exploited vulnerability need to be described.
- **Attack Technique.** A specific technique or technology to attack IT-Systems (Man in the Middle, Phishing, etc.). In particular the attacking actor, the conflict of interest and the exploited vulnerability need to be described.
- **Security Measures.** It describes a plan of sufficient measures to secure the intended interactions and to avoid the unintended interactions. Each security measure targets a vulnerability of the system in order to reduce a risk for a certain thread.

The *event view* contains all actual events through out the lifetime of the system. The event view instantiates the concept view of the system. It contains the following components:

- **Business Process.** The actual, running instance of the business model.
- **Accidents.** All actually happened accidents.
- **Attacks.** All actually happened attacks.
- **Security Operations.** Instances of security measures.

3.1.2 Procedure

The analysis procedure is an incremental and iterative process following the four views of the previously described model.

Step 1. World Analysis

At first, one has to outline the current state of the system under study. This includes description of:

- all **Assets** which must be protected
- all relevant **IT-Systems**
- all involved **Humans** and their **Conflicts of Interests**
- all known **Vulnerabilities**

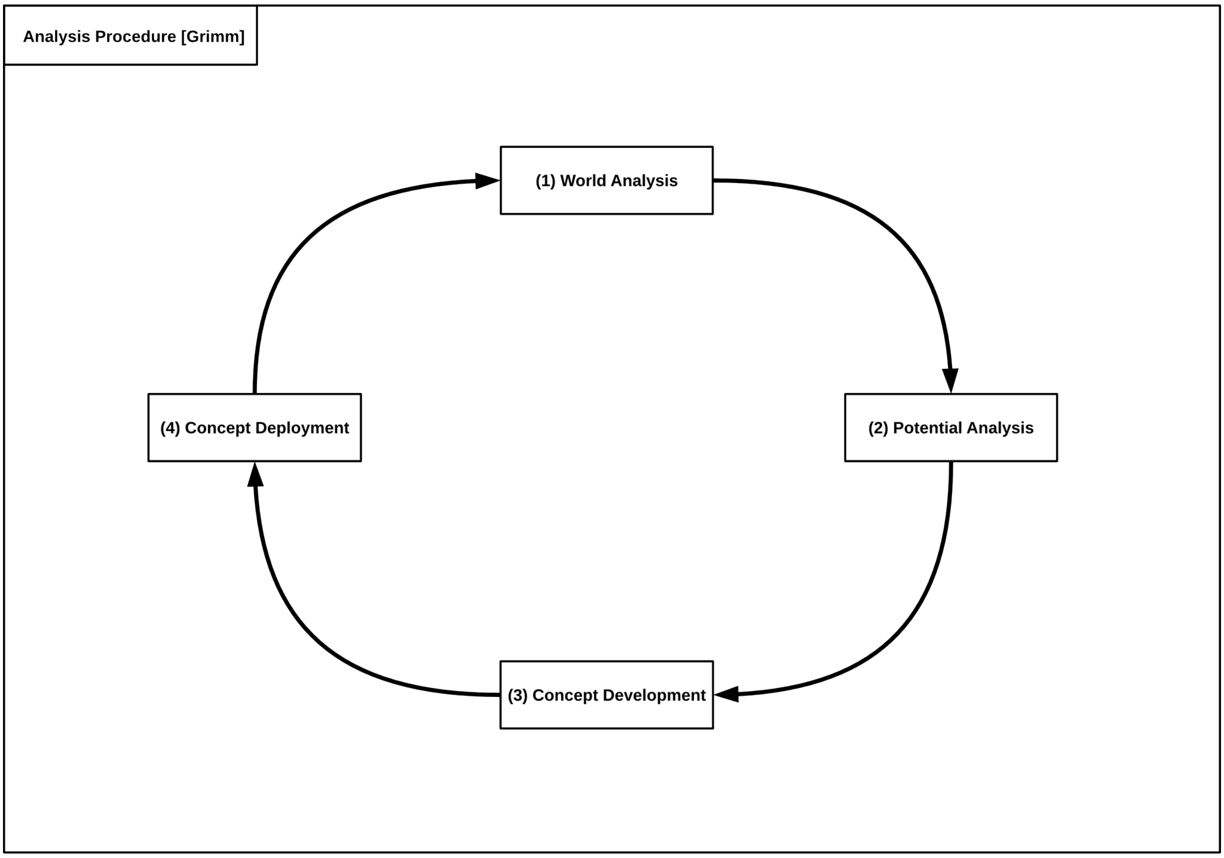


Figure 4: The IT Security Reference Analysis (Grimm)

- and all important **Interactions** between the former components.

Step 2. Potential Analysis

Secondly, one needs to outline the potential interactions of the system under study. This includes both the unintended interactions (threads) and the intended interactions (business objectives). This step produces four artifacts:

- a *threat specification*, which identifies the threat, its targeted assets, the involved actors and their conflicts of interest.
- a *threat risk evaluation*, which quantifies the likelihood of a threat manifestation in relation to its associated loss.
- a *security requirement specification*, which specifies requirements in order to deal with identified hazards

Step 3. Concept Development

Based on **Step 2.**, the identified hazards are used alongside realistic accident scenarios and attack techniques to create a *risk matrix*. With this matrix it is possible to decide if the risk is acceptable or not. Together with the previously specified security requirements, the matrix is used to define adequate security measures. Like a business model is an abstract concept to achieve business objectives, this step creates an concept to improve the system's security.

Step 4. Concept Deployment

Finally, the security measures have to be implemented. Additionally, all business operations, accidents, attacks and executed security operations will be recorded in the following time.

The implementation of security measures changes the world view (e.g. IT-systems, actors) and renders the conducted analysis outdated. So this analysis procedure needs to be conducted again.

3.1.3 Abstraction Levels of the Reference Model

The Reference Model can be used on different levels of abstraction. This means each component can be used within a wide range of granularity, for instance the security measure *Encryption* can be explored in general or on the level of different concrete encryption tools; or on the even finer level of concrete algorithms.

The utilized abstraction level is not important for the analysis procedure, it depends on the intended audience for the analysis. However, it is important to use one abstraction level consistently through out the analysis.

3.2 Live+Gov Privacy Protection Analysis

3.2.1 Step 1. World Analysis

3.2.1.1 Assets: Privacy

In this document we focus our attention to only one asset: The privacy of the citizen.

Our definition of privacy is described in detail in Chapter 2. In Section 2.2 we define privacy as the “control over private data” and introduce the following seven different types of privacy:

1. Privacy of the Person
2. Privacy of Behaviour and Action
3. Privacy of Communication
4. Privacy of Data and Image
5. Privacy of Thoughts and Feelings
6. Privacy of Location and Space
7. Privacy os Association

We refer to Section 2.2 for more details.

3.2.1.2 IT-Systems & Interactions

This section outlines the general architecture (Figure 5) of IT systems for public monitoring comparable to the Live+Gov project. This includes a description of it technical infrastructure and the interactions between its components.

The IT infrastructure of the Live+Gov system, i.e. the Live+Gov toolkit and the customization of the software components are described in detail in various project deliverables: D4.1, D4.3, D1.1, D5.1. In this section we give an abstraction of those systems from the perspective of WP1.

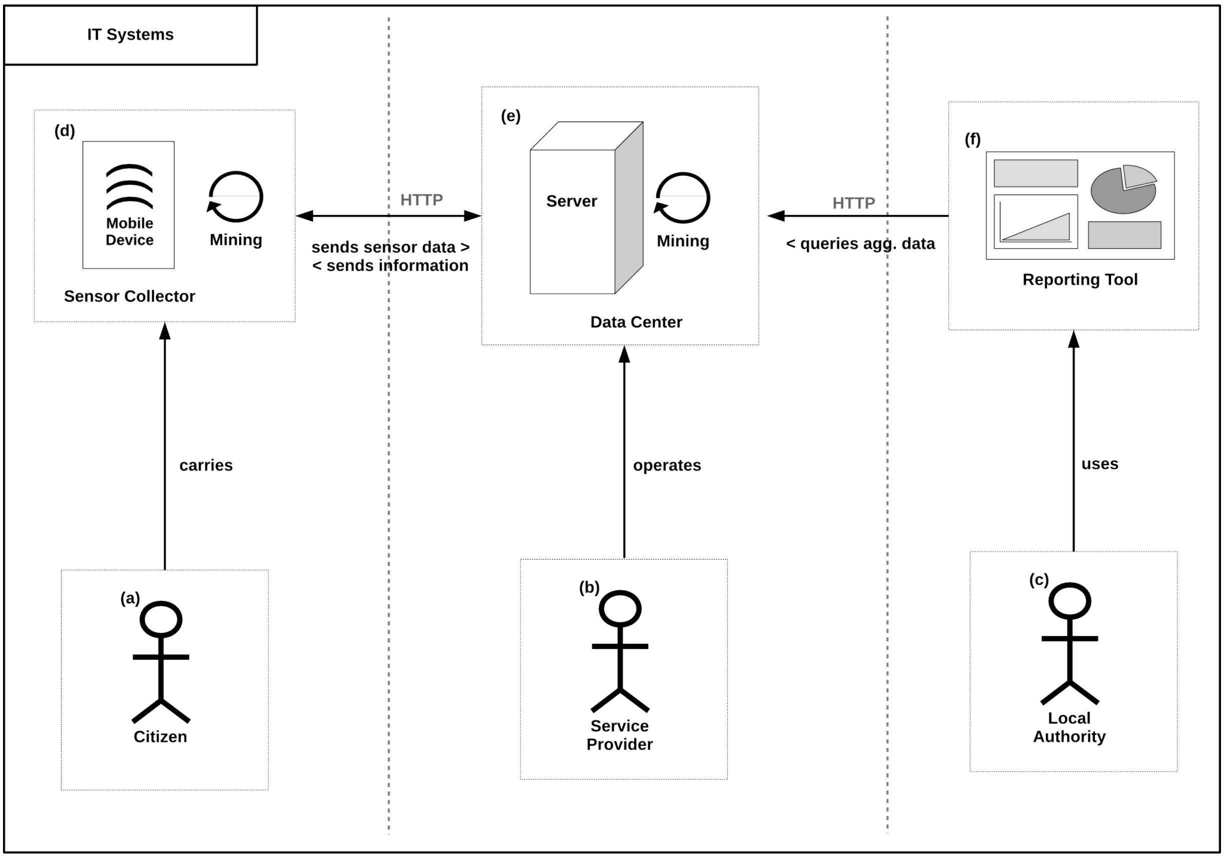
A *citizen* (a) carries a mobile device running the *sensor collector* application (d). The sensor collector application is able to collect various kinds of sensor data (accelerometer, GPS, GSM, ...) and can sent the raw data back to the *data center* (c). The sensor collector can also be able to perform certain data mining operations.

Examples of such data mining operations include human activity recognition,the detection of service lines, detection of characters and faces from images.

In the Live+Gov Project mobile devices are used in particular for the following activities:

1. collection of GPS samples,
2. mining of Human activities (HAR) based on accelerometer samples,
3. collection of reports consisting of an image, free text and selected categories.

The *data center* stores and processes sensor data collected with the sensor collector application. It can also take into account data obtained from third parties, like the current positions of trains.



Legend:

- **(a) Citizen:** User of the L+G client application whose privacy is at stake.
- **(b) Service Provider:** Provides technical infrastructure.
- **(c) Local Authority:** Provider of the L+G system.
- **(d) Mobile Device:** Runs the L+G client application, produces and stores data sensitive to the users privacy.
- **(e) Data Center:** Runs the L+G services, processes and stores user data.
- **(f) Report Tool:** Interface to aggregated user data.

Figure 5: IT Systems

The data center can sent mining end products (traffic jam reports, bus schedule) back to the mobile device of the citizen.

In the Live+Gov Project data centers, in particular, perform the following activities:

1. storage of login credentials, name, and email address for each citizen,
2. storage of collected GPS samples (user id, timestamp, GPS location),
3. storage of HAR results (user id, timestamp, HAR result),
4. detection of service lines based received GPS samples (SLD),
5. storage of SLD results (userid, timestamp, SLD result),
6. sent back SLD results to mobile device,
7. storage of received reports (user id, timestamp, report),
8. detection of inherent patterns of the received reports.

The *Service Provider* (b) provides and operates technical infrastructure like the Data Center and the *Reporting Tool* (f). The Reporting Tool queries the Data Center for aggregated data to visualize in form of charts and other means suitable to help understanding of monitored citizens.

Local Authorities (c) use the reporting tool to get information in order to understand citizen movement and improve public services. In such systems, the most valuable information for local authorities is not the raw sensor data, but aggregated views on the mining end products.

In the Live+Gov Project the reporting tools allows, in particular the following queries:

1. show aggregate information about which routes citizens take starting from a given location,
2. show the average waiting time of a citizen for each bus stop,
3. show routes where citizens where running to catch a bus,
4. show locations of all reports in a given time window,
5. visualize detected patterns in the report data,
6. view individual reports.

3.2.1.3 Actors

This section describes the human actors previously introduced within the IT system architecture (3.2.1.2). Also the additional actor *External* is introduced as a person with no special access rights.

Citizen. Citizens are persons who use the mobile device as users of the provided software. Their main motivation for using the software is to gain a higher level of convenience in their daily activities. For example they may have access to real time bus schedules or reports about traffic jams, that help them to avoid long waiting times. Also they generally benefit from improvements of public infrastructure by local authorities, which is triggered by issue reports.

By using the application citizens are sharing personal information like name and address, as well as data gathered from mobile sensor with the service provider. This data can be exploited in ways that are harmful to the citizen. This need to protect this protection is manifest in several laws and constitutions as the right to privacy and data protection. The citizen has a vital interest in having his legal rights protected and enforced.

If the right to privacy is violated, there is a magnitude of potential harms that interfere with other interests of the citizen. This includes just annoying spam, where their technical identity is used to send unwanted commercials. More severe phishing attacks can exploit personal information, and try to manipulate citizens into disclosing credentials like TAN numbers and disrespect their financial interests. Information about the current location (GPS) or frequently used routes (stalking) can be used to attack and directly harm the health of the citizen. Information about medical conditions inferred by sensor data (e.g. fintness trackers) are of interest to insurance companies, which may affect the pricing of policies and can interfere with financial interests of the citizen.

Also it is well established (cf. [3]) that the very act of being monitored can have impact on mental health and performance, promotes distrust and breeds conformity.

In short, citizens are interested in

- physical wellbeing and health
- financial profit
- convenience
- legitimate use of personal data
- non-disclosure of personal data to peers of the citizen
- not being monitored.

External. Externals are persons who do not have privileged access to the IT systems, and are willing to break laws, security constrains and norms in order to promote their interests.

If the external is in some kind of relationship to the citizen, like a friendship or business partnership, the external can have a direct interest in gaining information about the citizen in order to increase their power.

Another common interest of an external is financial profit. For example they want to obtain access to critical systems to steal sensitive data or to get the system under their control. Controlled systems could be leased as part of a bot net. Stolen data could simply be sold as is or used for illegitimate purposes, e.g. spam or phishing attacks - or excessive data mining.

Because local authorities are involved in the general outline of the Live+Gov system, the possibility for politically motivated attacks is given.

Externals could want to harm or destroy the systems in order to damage the reputation of local authorities (politicians or other officials) or to make a political statement of their own.

Another possible motivation for external activities could be social appreciation. A hacker could attack critical infrastructure just to prove his skills.

In short, externals are interested in:

- increase power over citizen
- financial profit
- political activism
- social standing.

Service Provider. Service Providers operate the technical infrastructure (hardware and software) of the IT System. They are private companies and legal persons in their own right, but also employ a number of people with diverging interests, including: administrators, who maintain and operate the running system; programmer/developer, who develop the system; a support manager, who handles customer relations.

As companies, they are interested in gaining financial profit. Among other things, this depends on customer satisfaction, employee happiness and task complexity. Unsatisfied customers may not want to pay for the service or do not continue the business relation. Moreover, unsatisfied customers can create a bad reputation, which affects the market for future customers.

Customer satisfaction is connected with the quality of the offered product or service. This quality depends on the happiness of employees. Employees have a claim to professional excellence. They want to deliver a good job within their means. If employees cannot satisfy their demand for professional excellence, they might get discontent and deliver poor work. Moreover, unhappy employees can produce higher costs through sick days. The worst case scenario could be, that a discontent employee gets angry and steals data or harms the running systems.

At last, the financial success of service providers depends on the task complexity of the maintained infrastructure. The complexity of a task has to be in reasonable bounds, so that service providers can complete it within time, with a satisfying quality. If a task has a higher complexity than expected, financial loss is almost certain. Either Service Providers need to hire additional competence to meet schedule and requirements. Or service providers they stress the time-line, which also results in a higher man-hour salary ratio and additionally endangers customer satisfaction. Ultimately, high task complexities can affect employee happiness, if employees cannot complete it within their claim to professional excellence.

In short, Service Providers are interested in:

- financial profit
- good working conditions
- professional excellence
- manageable complexity.

Local Authority. Local authorities are public offices (ministry, agency, department, ...) or other external public entities which act as direct customers of service providers. They purchase a system specialized for their needs. For example a department for urban mobility, orders a system to better understand usage patterns and make improvement to the urban traffic flow.

Such systems are investments, and so naturally local authorities are interested in a profitable return, like increased ticket sales. However, the return of investment is not directly of financial nature. Like service providers their financial gain depends on customer satisfaction. Customers for local authorities are either citizens, who use their services, or politicians, who order their services. The satisfaction of both sides is interdependent.

Citizens are satisfied customers if the services, e.g. public mobility, work well. If citizens are happy, it is more likely that politicians gain reputation, as they organize the public services through local authorities.

The first important step of improving the public services is by obtaining business intelligence. For the urban mobility scenario the Live+Gov systems provide insight in form of traffic jam detection and usage pattern mining, which allow local authorities to focus their efforts to the most important sites.

Additionally, since local authorities act like corporations comparable to service providers, they are also interested in good working conditions. Discontent employees may harm the system by e.g. disclosure of privacy sensitive data.

In short, local authorities are interested in:

- financial profit
- political reputation
- business intelligence
- good working conditions.

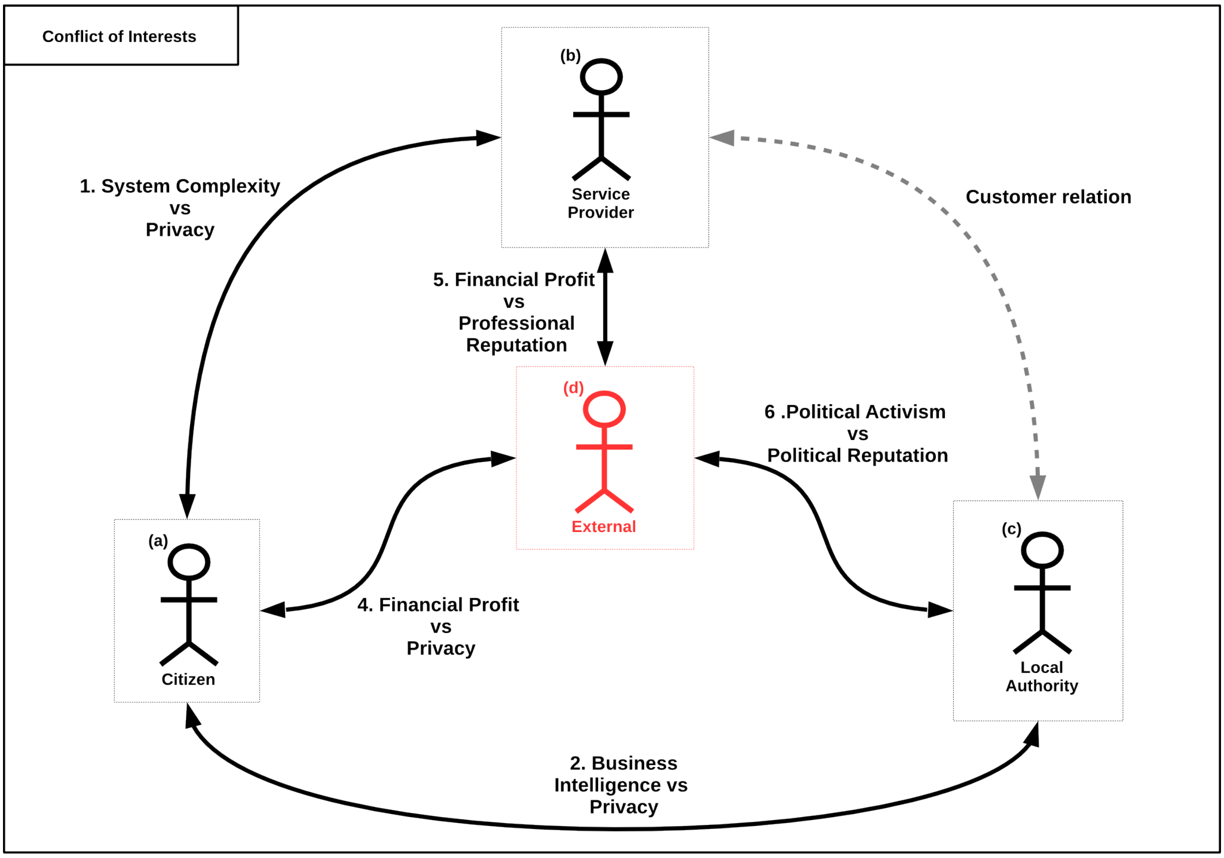

Conflict of Interests
(b)
Service Provider
1. System Complexity vs Privacy
Customer relation
5. Financial Profit vs Professional Reputation
(d)
External
6 .Political Activism vs Political Reputation
(a)
Citizen
4. Financial Profit vs Privacy
(c)
Local Authority
2. Business Intelligence vs Privacy

Figure 6: Live+Gov Conflicts of Interests

3.2.1.4 Conflicts of Interests

This section outlines the Conflicts of Interests (Figure 6) between the actors of the proposed IT system architecture.

The individual interests of all actors is already described in the previous section and are not elaborated any further. The emphasis here is put on prominent existing conflicts, because they provide a foundation for vulnerabilities and subsequent threats.

1. System Complexity vs Privacy. Service Providers offer a service to Local Authorities, which is to provide and maintain a monitoring and mining system, e.g. for public mobility. This system shall produce business intelligence, so that Local Authorities can improve their public services. This task in itself has a high technical complexity and is the sole asset with financial return for Service Providers. However, this task operates on privacy sensitive data provided by monitored Citizens. In order to ensure their privacy, Service Provider would have to implement additional mechanisms, which allow Citizens to exercise control of their data. This will not only raise the complexity of the monitoring and mining system, Service Providers also have to layout the complexity in a comprehensible manner. To effectively enable Citizens to preserve their privacy, they need to know what happens with their data.

2. Business Intelligence vs Privacy. Local authorities order a monitoring and mining system from service providers, which allows them to produce business intelligence for public services. The system is an investment for local authorities, so they are interested in as much intelligence as possible to achieve a profitable return.

The gained intelligence is the result of data mining conducted on privacy sensitive data of participating citizens. They are interested in the successful usage of their data, in a sense that they are also benefactors, e.g. improvement of public mobility. The main interest of citizens lies in maintaining control over their data and protecting their rights to privacy. In order do that, they need full disclosure of the processing steps and the purposes their data is used for, and to be given a choice whether such processing should be allowed for their own data.

3. Power of External vs. Privacy (not shown in Figure 6). Externals which are in a social relation to the citizen can have an interest in obtaining further information in order to gain power. In the most simplistic example this could be somebody monitoring the activities of his marriage partner. Another example is a Government spying on it's citizens in order to suppress opposition. An additional twist in the last example is, that there can be legal regulations that require the service provider to support the Governments invasion of the citizens privacy.

4. Financial Profit of External vs Privacy. Externals can gain financial profit from stealing privacy sensitive data. For example by selling raw contact information to advertisers or by selling mined data to insurance companies, or intermediaries like scoring companies. In such cases, citizens lose complete control over their data.

5. Financial Profit of External vs Reputation of Service Providers. Externals have various business models as optional foundation for attacks on Service Providers. For instance, they can try to invade the infrastructure for e-espionage reasons, to get control over servers to create a bot-net or to steal user data. All these approaches are motivated by financial interests. Gathered information can be sold, zombie servers can be leased.

A successful attack proves the technical competence of service providers wrong and subsequently harms their professional reputation. This can lead to a loss of future customers or a decrease of stock price for registered companies. Eventually also the financial interests of service providers are endangered.

6. Political Activism vs Reputation of Local Authority. Besides monetary reasons, externals can be motivated by political reasons to attack the monitoring and mining system. Externals can break the system to make a political statement of their own, or they can steal user data to prove the system insecure. Both would harm the reputation of local authorities, who endangered the privacy of the citizens.

3.2.1.5 Vulnerabilities

This section outlines the vulnerabilities (Figure 7) of the proposed monitoring and mining system. Note that vulnerabilities are not necessarily of technical nature. The weaknesses of IT systems are often created due to misuse or misconfiguration of the various components by one or more actors.

Insecure Infrastructure. The proposed monitoring system consists of many hardware and software components, each with its own concrete weaknesses. For instance, operating systems can be outdated or not subject to frequent updates or virus scans. Web-applications can be carelessly implemented and not protected against SQL-Injections or Cross-Site-Scripting attacks. Databases can be ill-configured, so that access from outside the system is possible. All those weak points can be subject to various known exploit techniques.

Insecure Data Transmission. The proposed monitoring and mining system uses HTTP to

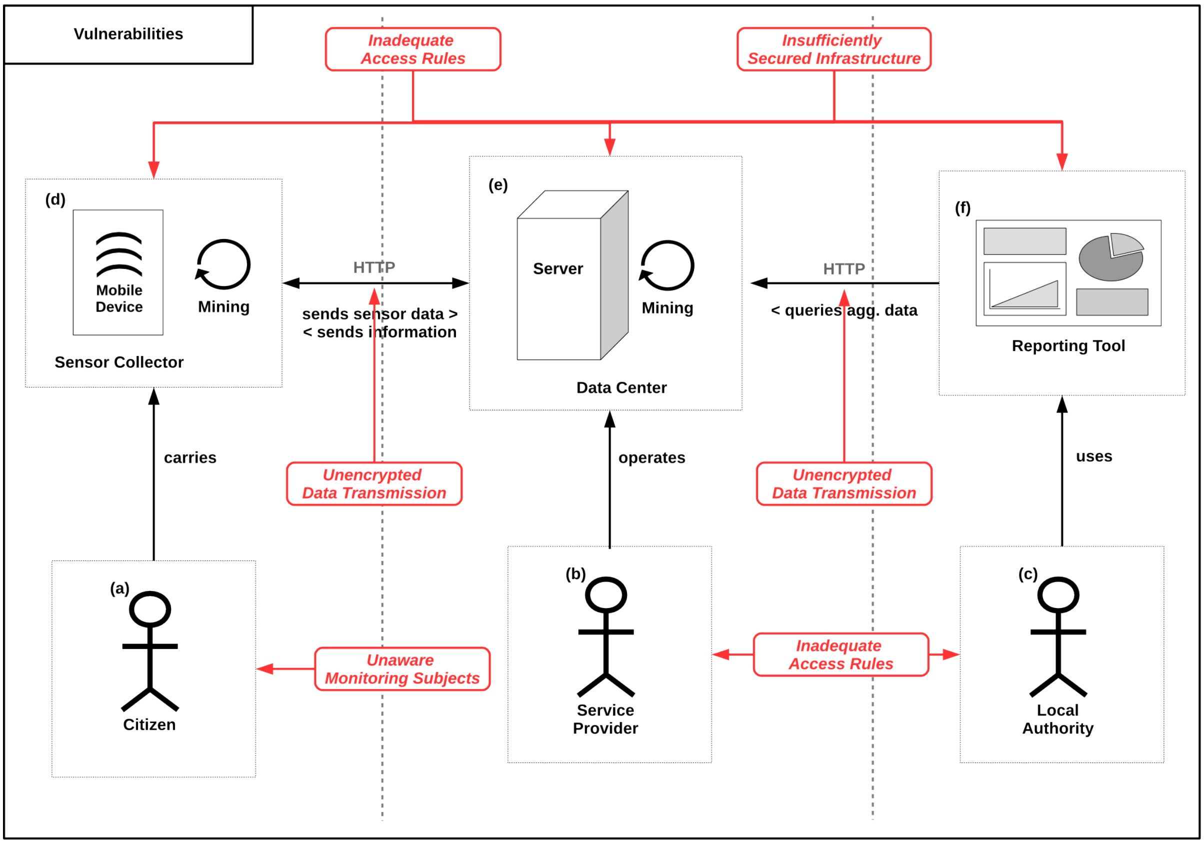


Figure 7: Live+Gov Vulnerabilities

exchange data between the Sensor Collector, the Data Center and the Report Tool. Per default, HTTP is a clear text protocol. This means, one can intercept the connection and read all sensitive information, which is send between the components. That is: passwords, raw sensor data and data mining results

Unhappy Employees. An Employee that is frustrated with his situation for a long time period constitutes a security vulnerability. On the one hand he might want to harm his employer directly, on the other he is increasingly susceptible for social engineering.

Inadequate Access Rules. The proposed IT system infrastructure has various accesses to privacy sensitive data. Service Provider staff has access to Data Center hardware and software like databases, web-servers and other inspection tools. Local Authority staff has access to the Report Tool. This all enables staff members to have potential access to privacy sensitive information. Those accesses have to be secured against unauthorized third parties. Moreover, we need to ensure that no single person has to many access rights. For example, a system administrator should not be able to secretly download the whole database on a flash-drive.

Unaware Monitoring Subjects. We define privacy as one's ability to control information about oneself. In order to do that, monitored subjects need to know, that they are monitored, who monitors them, what information is recorded and for what purposes. Subjects who are not aware of these things cannot effectively preserve control and thus lose their privacy. This vulnerability expresses itself as the lack of information material like a Privacy Policy including concise information about applied processing steps, access rules and disclosure to 3rd parties.

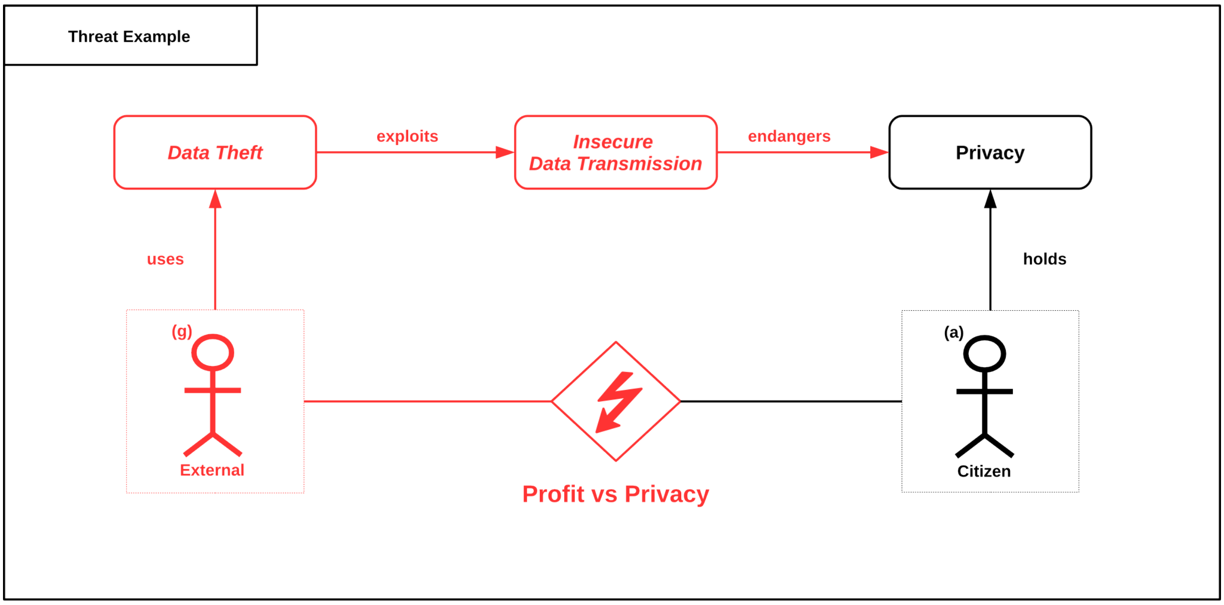


Figure 8: Threat Example

3.2.2 Step 2. Potential Analysis

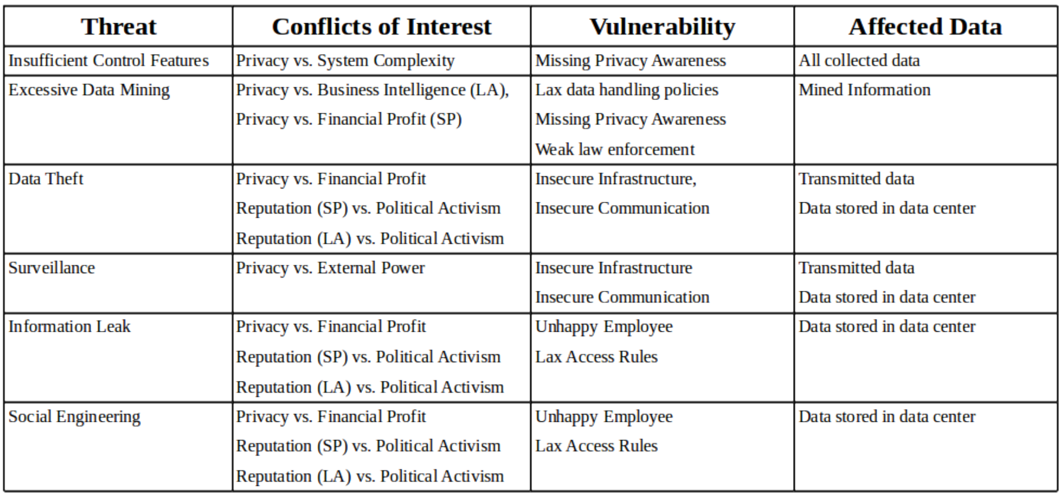
3.2.2.1 Threat Specification

Recall from the security model description in Section 3.1, that a threat is a potential interaction of the components that targets an asset. We restrict ourselves to the case of attacks, and the asset of privacy. An attack is an interaction that is executed by an actor in response to a conflict of interest by exploiting a vulnerability of the system. The alternative interaction of accidents are less relevant for us, since the violation of privacy always requires an actor that takes advantage of personal data.

In Figure 8 we illustrate the structure of threats at the example of “Data Theft”. Here the attack is executed by an external person, that harms the privacy of the citizen. He is motivated to do so by financial profit gained by selling personal information, which is in conflict with the citizens interest in his privacy. To get hold of the data the external exploits an insecured HTTP transfer of recorded data.

In this section we describe, in a similar fashion, threats for the citizen privacy in the Live+Gov system. This list can necessarily not be complete, but we make a best effort to cover the most relevant cases.

T1. Insufficient Control Features. The Service Provider does not offer tools for the Citizen to control his data. However, besides the missing features he provides a safe and secure system. This threat contradicts the definition of privacy as control over one’s information about oneself. As soon as collected data of Citizens is stored on Service Provider servers, all control over that data is lost. This is not necessarily due to bad intention, Service Providers may simply have forgotten to include such features during the development process. Although, control capabilities for Citizens add to the system complexity, which could motivate to omit those. Therefore interests of Citizens and Service Providers are in conflict, namely it is the Citizen’s *Privacy vs. System Complexity* for Service Providers. This is an abstract threat provoked by a general *Missing Privacy Awareness* in the minds of all actors in the Live+Gov context.



Threat	Conflicts of Interest	Vulnerability	Affected Data
Insufficient Control Features	Privacy vs. System Complexity	Missing Privacy Awareness	All collected data
Excessive Data Mining	Privacy vs. Business Intelligence (LA), Privacy vs. Financial Profit (SP)	Lax data handling policies Missing Privacy Awareness Weak law enforcement	Mined Information
Data Theft	Privacy vs. Financial Profit Reputation (SP) vs. Political Activism Reputation (LA) vs. Political Activism	Insecure Infrastructure, Insecure Communication	Transmitted data Data stored in data center
Surveillance	Privacy vs. External Power	Insecure Infrastructure Insecure Communication	Transmitted data Data stored in data center
Information Leak	Privacy vs. Financial Profit Reputation (SP) vs. Political Activism Reputation (LA) vs. Political Activism	Unhappy Employee Lax Access Rules	Data stored in data center
Social Engineering	Privacy vs. Financial Profit Reputation (SP) vs. Political Activism Reputation (LA) vs. Political Activism	Unhappy Employee Lax Access Rules	Data stored in data center

Figure 9: Threats

T2. Excessive Data Mining. The Service Provider and/or the Local Authority secretly extract more private information from the collected data, than the Citizen agreed to. But results of the mining process create no disadvantages for Citizens, because there is no disclosure to third parties. This could be the case for a Service Provider, who wants to test a new product and uses the pre-existing data collection. Or for a Local Authority, who wants to analyze the data collection regarding fare evasion. However, the Citizen has not agreed to such data processing nor could he, since it is conducted secretly. This disables a Citizen to control his data adequately. Thus there are two possible conflicts: *Privacy vs. Financial Profit of Serivce Provider* and *Privacy vs. Buisness Intelligence of Local Authority* The threat can be provoked by either lax data handling policies of both Service Providers and Local Authorities, or a weak law enforcement of existing supervision. But the main issue, which can lead to such threats, is again a general *Missing Privacy Awareness*.

T3. Data Theft. An External infiltrates infrastructure in order to steal personal data and sell it on the black market. Also the External might be motivated politically and wants to harm the reputation of the Service Provider or the Local Authority. Anyways, this threat would be manifested through a technical attack on either hardware (Packet Capture) or software (buffer overflow, SQL injection). Such a successful attack could harm the reputation of both Service Provider and Local Authority. The technical competence of Service Providers would be proven wrong, therefore they would lose professional reputation. Local Authorities would lose their political reputation, as it would seem like they had endangered data of Citizens, which lose complete control. Thus this threat is defined by three conflicts: *Privacy vs. Financial Profit*, *Reputation of Service Provider vs. Political Activism* and *Reputation of Local Authority vs. Political Activism*. Also this threat describes the classical scenario, where attacks are provoked by *Insecure Infrastructure* (SQL injection) and *Insecure Communication* (Packet Capture).

T4. Surveillance. An External infiltrates infrastructure in order to obtain information about the citizen and exploit it directly. In this scenario the external is supposed to have some direct relationship to the citizen which motivates his interest to obtain personal information. Examples could be a public institution that wants to gain information about planned activities of the citizens (e.g. Nixon's Watergate scandal or the recent prosecution of Guardian journalists by GHCQ). Another example is an insurance company that seeks to get information about the citizens lifestyle in relation to the insured risk, like car accidents or health hazards.

In this threat the privacy interest of the citizen is in conflict with the aspirations for power over the citizen by the externals.

T5. Information Leak. Like an external person the Data Theft Scenario an employee of the service provider or the Local Authority has selfish interests to gain money, make political statements or harm his employer. In order to pursue this interest he can steal personal data and sell it or release it to the public. The corresponding conflicts of interests are: *Privacy vs. Financial Profit* of the Employee, *Reputation of Service Provider vs. Political Activism* of the Employee and *Reputation of Local Authority vs. Political Activism* of the Employee. The vulnerability constitutes of the existence of *Unhappy employees* itself and possibly *lax access rules* that enable the employee to obtain large amounts of data unnoticed.

T6. Social Engineering. This scenario an external manipulates an employee of a Service Provider or the Local Authority to leak information to the external person. It is thus combination of the Data Theft and Information Leak scenario. The conflicts of interest are *Privacy vs. Financial Profit* of the External, *Reputation of Service Provider vs. Political Activism* of the external and *Reputation of Local Authority vs. Political Activism* of the external. The exploited vulnerabilities are, again, the existence of *Unhappy employees* and possibly *lax access rules* that enable the employee to obtain large amounts of data unnoticed.

3.2.2.2 Threat Risk Evaluation

In this section we will associate to every identified threat a corresponding risk. Recall from 3.1 that a risk is the expected loss that is associated to the threat. Therefore, we have to quantify the likeliness of the threat to occure and the harm or loss done in this case. The quantification of likeliness will be solely based on rough judgment of the authors. The quantification of loss, will be made in a two step process. For each threat listed in Figure 9, we have analyzed the affected personal data of the citizen. For each possible data type (e.g. GPS) we analyze the impact on the seven different types of privacy in Section 2.2.4. In combination we can quantify roughly the impact of each threat on the citizens privacy. Both evaluations are necessarily fraught with a high level of uncertainty.

For the quantification of the loss in case of a threat scenario we use the following rough calibration:

- 3: High. Leak of information to peers (e.g. public) which impacts citizen.
- 2: Medium. Undisclosed processing of personal data or disclosure to third parties that are unrelated to subject.
- 1: Low. Loss of control over data.
- 0: None

For the quantification of likeliness the following scale is used:

- 4: Always.

- 3: High. Occurs once in 10 cases
- 2: Medium. Occurs once in 100 cases
- 1: Low. Occurs once in 1 million cases
- 0: Impossible

The quantification of the risk, we add the values for loss and likeliness of the corresponding threats. Note, that loss and likeliness scales have a logarithmic character, so that that addition of those scales corresponds to multiplication of the usual scales.

The likeliness, loss and the resulting risks assigned to the threats are discussed in the following paragraphs and summarized in Figure 10.

T1. Insufficient Control Features. The occurrence of this threat is dependent on the design on the system and given in our case, since we do not give the citizen control over his data once it is recorded. Therefore the Likeliness is evaluated as 4 (Always). The associated, risk is 1 Low on our scale, since no direct harm is done to the citizen by exploiting the data.

Hence the resulting risk is calculated as $4+1=5$.

T2. Excessive Data Mining. We assess the likeliness of excessive data mining to be 3 (High), since these kind of analysis can be performed within the walls of the service provider, without somebody else noticing, and the service provider himself has an interest in this activity.

The associated loss, on the other hand can be substantial, i.e. 2 (Medium). For example when GPS data is linked to data from telephone books the identity of the citizen can be revealed and personal details like visits to doctors. In the threat scenario, this information is not leaked to third parties, (which would justify an even higher loss assessment), but the very existence of this information violates the citizens privacy.

Hence the resulting risk is calculated as $3+2=5$.

T3. Data Theft. The likeliness of a targeted attack by a third party is dependent on the popularity of the offered service and financial value of the captured information. Moreover, the amount of manual work required to infiltrate a custom build system is significantly higher that that of compromising a standard software solution. In the scenario we assume a moderate popularity in a single metropolitan area, with around 10.000 users and storage of data of only limited financial value (no addresses, no payment information). Therefore the likeliness assessment is $1-2$ (Low-Medium).

The harm of leaked information to a criminal party is 3 (High). Hence the resulting risk is calculated as $4-5$.

T4. Surveillance. In the surveillance scenario an party related to the citizen, like a company where he is customer of, or a government agency, seeks to obtain sensitive information from our service.

The likeliness of such an intrusion is hard to asses, and depends again on the popularity of the service. If a high popularity is reached we have recently learned that spying by government agencies is very likely to occur. The barrier for companies that do not operate the infrastructure used to transmit the data a surveillance attack is however very hard to perform. Therefore we assess the likeliness of the threat with $1-2$ (Low-Medium).

Threat	Likeliness	Loss	Risk	Recommendation
T1. Insufficient Control Features	4	1	5	R1, R2
T2. Excessive Data Mining	3	2	5	R3, R4, R5
T3. Data Theft	1 - 2	3	4 - 5	R6
T4. Surveillance	1 - 2	3	4 - 5	R7
T5. Information Leak	1	3	4	R8
T6. Social Engineering	1	3	4	R8

Figure 10: Live+Gov Risk Evaluation and Recommendations

The harm of leaked information to a related party is 3 (High). Hence the resulting risk is calculated as 4 – 5.

T5/6. Information Leak and Social Engineering.

In our scenario we assume that the culture and ethics inside the service provider company and local authority are very high, so that the information leak scenario has a likeliness of 1 (Low).

The harm of such an information leaked is 3 (High), so that the resulting risk is calculated as 4.

3.2.2.3 Privacy Recommendations

In the preceding section we have identified the main risks for the users privacy. In this section we derive recommendations or requirements for a system that addresses these risks. Some of these requirements are implemented as security measures in our systems and discussed in the following chapter 4.

In order to address the threat with the highers risk, Insufficient Control (T1) of the citizen, we need to give the citizen back the control over its data inside the system. The most direct way to do this is to provide a web-based *Privacy Dashboard (R1)* which allows the citizen to view, edit and delete all information about his person that is stored inside the system. Also control applied processing and disclosure of the data to third parties should be given to the user, at least in the form of an opt-out or veto option.

A necessary pre-requirement for effective control of the citizen over his data is information and comprehension of the intended data capturing and processing steps. Therefore a *Privacy Policy (R2)* that is easily readable and contains all important information is essential. Moreover, the existence of a Privacy Policy is a legal requirement (cf. Section 2.1.3).

The threat with the second largest risk is (T2) Excessive Data Mining. Contrary to common belief, it is neither legal nor ethical to process personal data for by new methods or for new purposes that were not stated and explained to the citizen at the time of data collection. Also the common practice of obtaining far-reaching permissions from the citizens inside the privacy policy is neither an ethical or legal solution to the problem (cf. Art. 6 in Section 2.1.3).

To address this threat awareness about the limitations of data processors inside the company is a key element. As one mean to establish such a culture of privacy respect, we recommend to prepare an document called *Data Handling Guidelines (R3)* intended for internal use that explains the concrete processing steps and purposes that are permitted by the citizens. This guideline should also be structured in a way that it covers the legal notification requirement from the EU Data Protection Directive (cf. Section 2.1.3). In particular the following information should be provided for each processing task: The name of the controller, purpose of processing, description of the data categories, recipients of the data if disclosed, transfer to third countries and a description of security of processing.

If further processing should be performed, it is necessary to seek additional permissions from the citizen. A simple email explaining the planned processing steps, and *asking for permission (R4)* would be enough for this purpose. The permission can be given via an embedded link that shall be followed in order to signal agreement.

An alternative measure to address the risk of excessive data mining is the *anonymization (R5)* of data. When all direct- or indirect links to the identity of the person are removed, no violation of the citizens privacy caused by arbitrary processing. However removing all such links is a challenging tasks, and full anonymity is often not achieved, cf. [13].

The protection from threat scenario (T3) Data Theft is a case of classical *IT infrastructure security (R6)*. The storage and processing infrastructure has to be secured using firewalls, up-to data software versions and proper authentication mechanisms.

The protection from threat scenario (T4) Surveillance focuses on the communication channels. They are target of wiretapping attacks by intermediaries or externals with access to the communication infrastructure. Strong *encryption (R7)* should be used to make it harder for externals to read the content of the transmitted data.

Threat scenarios (T5) Information Leak and (T6) Social Engineering target the vulnerability of unhappy employees. Therefore a trustful, *healthy company culture (R8)* should be maintained.

In summary we recommend the following measures to secure the citizens privacy:

R1 Privacy Dashboard. A tool which allows the citizen to view, edit and delete all data personal data that is stored in the system.

R2 Privacy Policy. A document, that informs the citizen about the collection and processing of personal information. It should at least contain the legally required information.

R3 Issue Data Handling Guidelines that explain the permitted processing methods and purposes.

R4 Ask the citizens for permissions before applying further processing via Email.

R5 Anonymize personal data before processing.

R6 Securing of Storage and Processing infrastructure using e.g. firewalls.

R7 Securing communication channels using encryption.

R8 Maintain a healthy, trustful relationship with your employees.

The mapping of these recommendations to the threats is summarized in Figure 10.

4 Privacy Protection Measures in the Reality Sensing Infrastructure

In this chapter we explain the measures we have taken in order to protect the citizens privacy. We have used the recommendations derived from the Security Analysis performed in Chapter 3 as a guideline.

4.1 Privacy Dashboard

The Live+Gov Privacy Dashboard is a software component that allows the citizens to view and control the data about themselves that is stored in the Sensor Data Storage server. It implements privacy recommendation R1 of Section 3.2.2.3. The foundation of the Privacy Dashboard component is the inspection tool that was introduced in D1.1 as a means to perform basic quality controlling of the collected data.

The Privacy Dashboard is currently endowed with the following views:

- **Login Screen** (cf. Figure 11). Before access to the dashboard is granted, the user is prompted for credentials. The credentials are generated by the sensor collection application on the mobile device. The username can be freely chosen by the user. The password is then generated by the mobile application and transferred to the server alongside with the recording.

- **Recording Overview** (Cf. Figure 12). This view shows all recordings that the user submitted to the Data Storage Service including basic meta information like start data of recording, duration. The user can click on each recording in order to view the recorded data in the Raw Data View.

  An important feature of the recording overview table is the delete button (cross on the very right.) That allows the user to selectively delete recordings, that shall not be analyzed by the Live+Gov system. Internally the click on the delete button only flags the corresponding trip for deletion. The actual deletion is carried out as a batch job on a regular bases. This addition step prevents accidental loss of data.

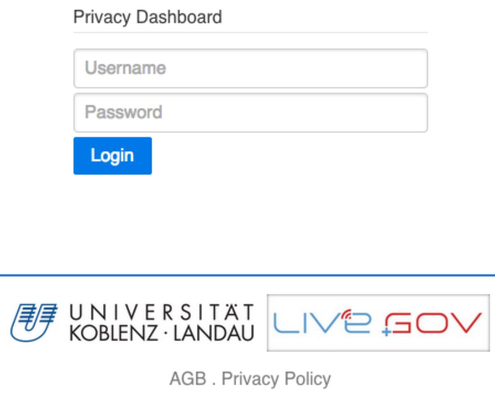
- **Raw Data View** (cf. Figure 13). The raw data view shows all recorded raw data to the user. GPS data is visualized on a map. Motion sensor data is displayed as plots. Wifi and GSM data is displayed in lists. The raw data from each sensor can be downloaded individually as CSV file.

- **Processing View** (cf. Figure 14). The processing view shows the data mining end-products for each processing step. The Figure shows the processing view for the Activity Recognition. On the top a time-line is show that visualizes the different recognized activities via color-coding. A map below displays the corresponding GPS track with the same color coding.

  A similar view for Service Line Detection results is currently under development.

- **Privacy Settings.** This view contains the privacy policy shown in Section 4.2.

  We intend to include settings concerning the expiration date of the recordings, agreements to certain processing steps, view logs of performed processing.



Privacy Dashboard

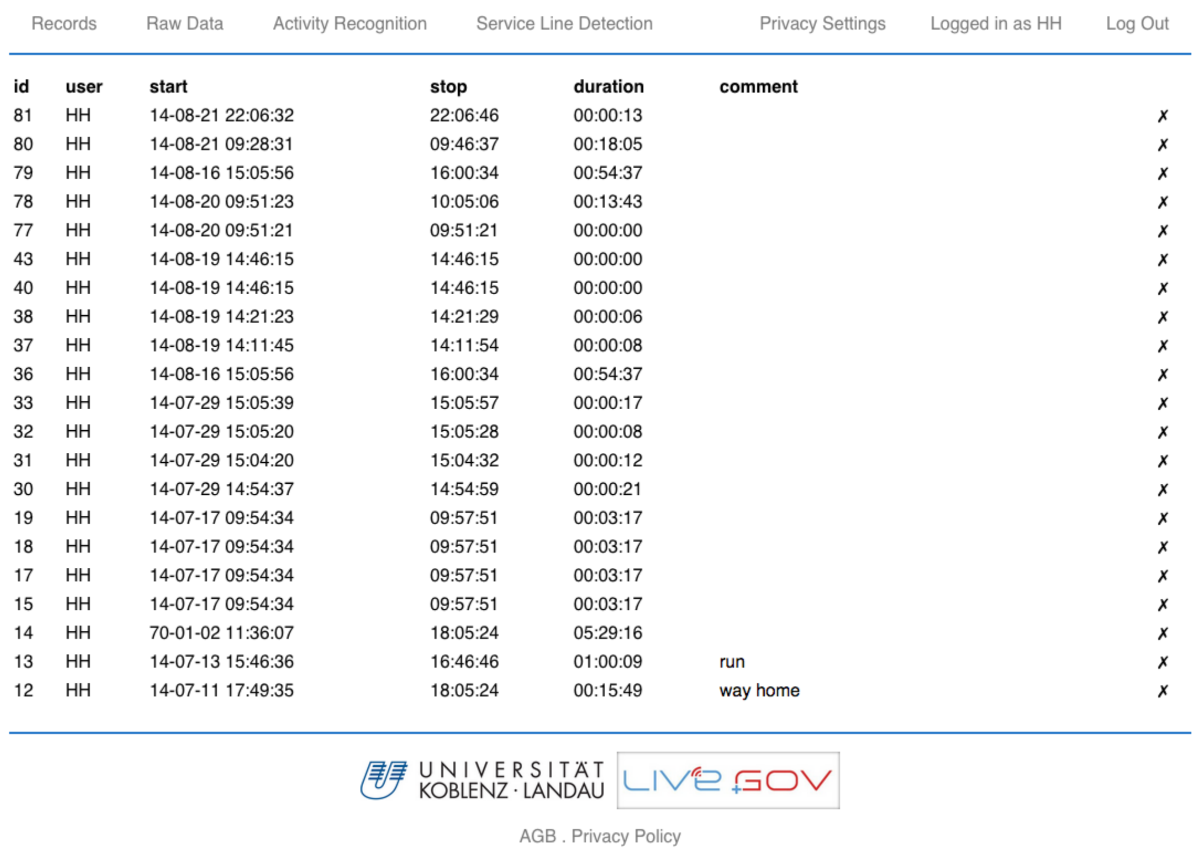
Figure 11: Live+Gov Privacy Dashboard Login

The Privacy Dashboard is implemented on a nodes.js[2] web server using Express.js[3] as a web-framework and d3.js[4] as plotting library. The data is pulled directly from the PostgreSQL database of the Data Storage Service.

[2] http://nodejs.org

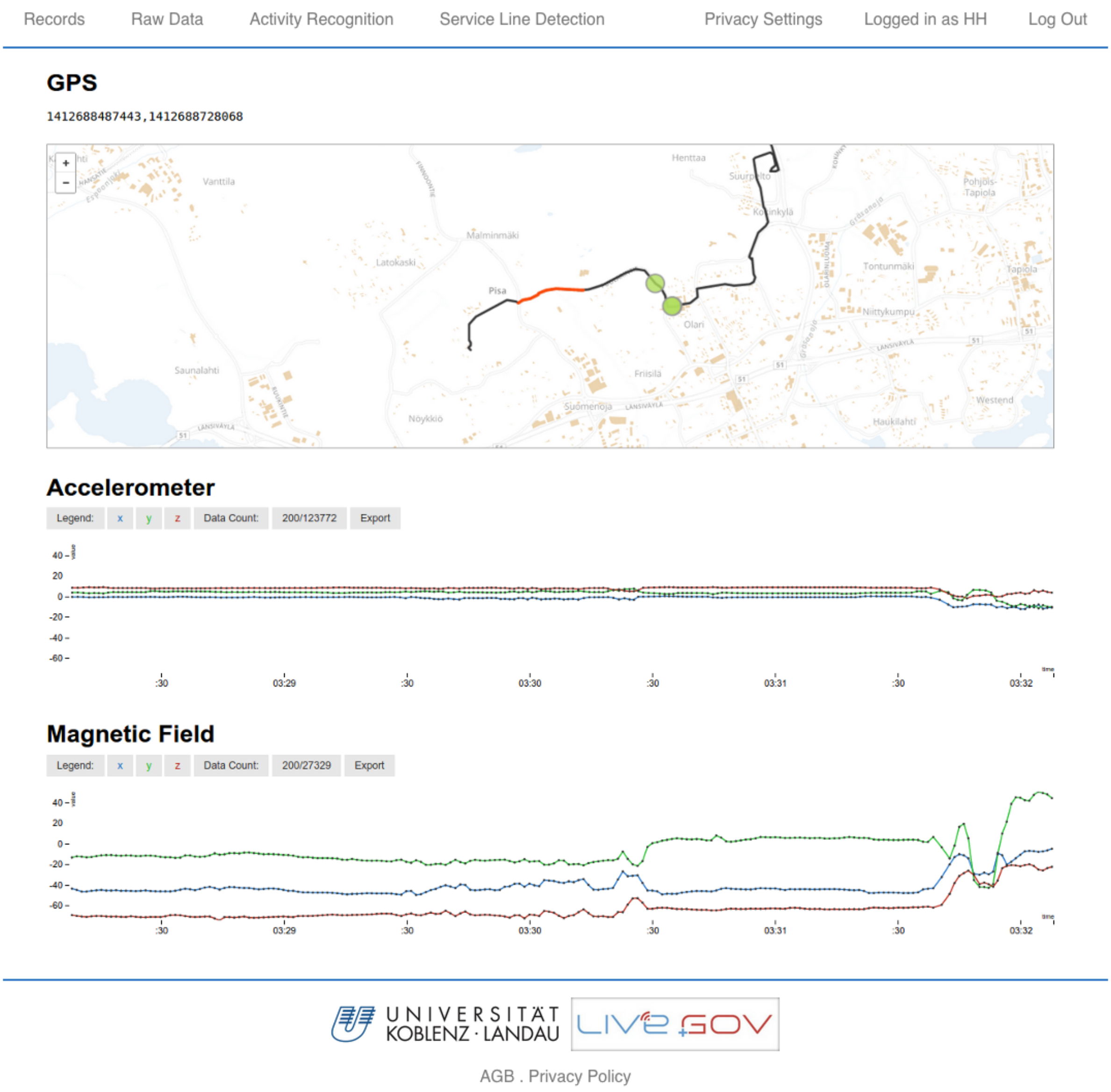
[3] http://expressjs.com/

[4] http://d3js.org/



id	user	start	stop	duration	comment	
81	HH	14-08-21 22:06:32	22:06:46	00:00:13		✗
80	HH	14-08-21 09:28:31	09:46:37	00:18:05		✗
79	HH	14-08-16 15:05:56	16:00:34	00:54:37		✗
78	HH	14-08-20 09:51:23	10:05:06	00:13:43		✗
77	HH	14-08-20 09:51:21	09:51:21	00:00:00		✗
43	HH	14-08-19 14:46:15	14:46:15	00:00:00		✗
40	HH	14-08-19 14:46:15	14:46:15	00:00:00		✗
38	HH	14-08-19 14:21:23	14:21:29	00:00:06		✗
37	HH	14-08-19 14:11:45	14:11:54	00:00:08		✗
36	HH	14-08-16 15:05:56	16:00:34	00:54:37		✗
33	HH	14-07-29 15:05:39	15:05:57	00:00:17		✗
32	HH	14-07-29 15:05:20	15:05:28	00:00:08		✗
31	HH	14-07-29 15:04:20	15:04:32	00:00:12		✗
30	HH	14-07-29 14:54:37	14:54:59	00:00:21		✗
19	HH	14-07-17 09:54:34	09:57:51	00:03:17		✗
18	HH	14-07-17 09:54:34	09:57:51	00:03:17		✗
17	HH	14-07-17 09:54:34	09:57:51	00:03:17		✗
15	HH	14-07-17 09:54:34	09:57:51	00:03:17		✗
14	HH	70-01-02 11:36:07	18:05:24	05:29:16		✗
13	HH	14-07-13 15:46:36	16:46:46	01:00:09	run	✗
12	HH	14-07-11 17:49:35	18:05:24	00:15:49	way home	✗

Figure 12: Live+Gov Privacy Dashboard Recording Overview



GPS

1412688487443,1412688728068

Accelerometer

Legend: x y z Data Count: 200/123772 Export

Magnetic Field

Legend: x y z Data Count: 200/27329 Export

AGB . Privacy Policy

Figure 13: Live+Gov Privacy Dashboard Raw Data View

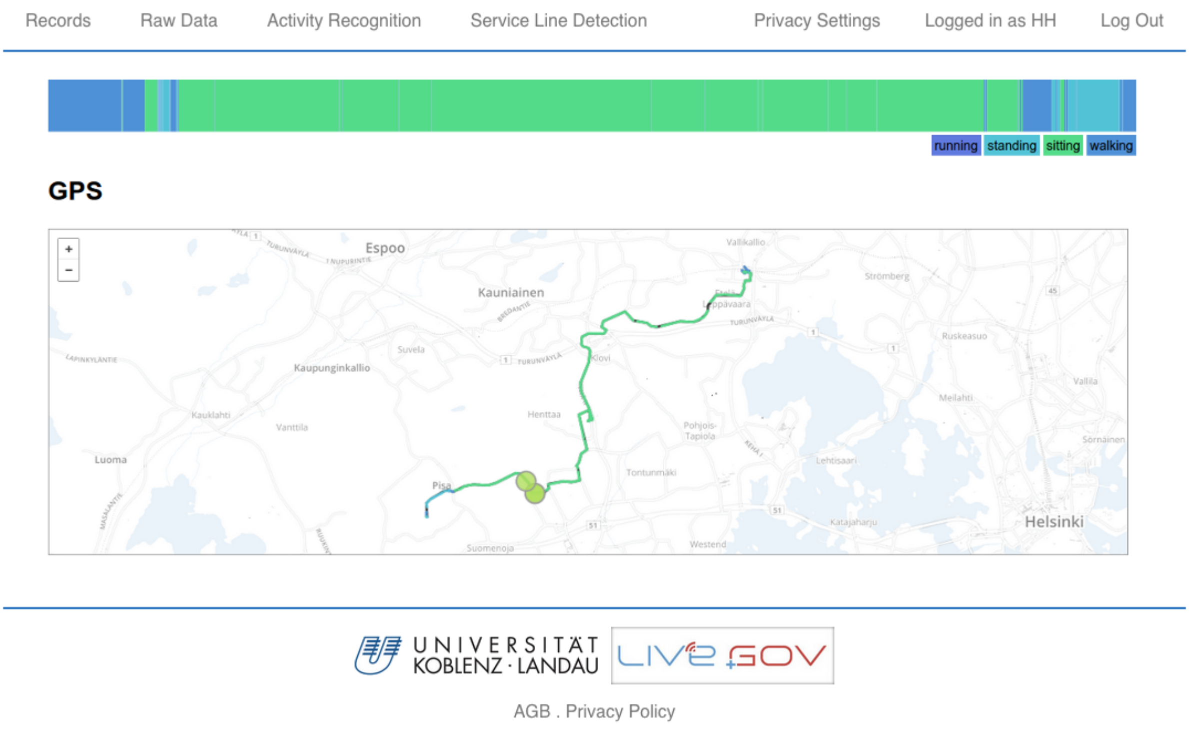


Figure 14: Live+Gov Privacy Dashboard Processed Data View

4.2 Privacy Policy

Implementing Privacy Recommendation R1 of Section 3.2.2.3 and following legal requirements of the Data Protection Directive (cf. Section 2.1.3) we include the following Privacy Policy on our Mobile Application and Website.

Privacy Policy
==============

What Personal Information do we Collect?
The Mobile Sensor Collection Application collects the following data from your mobile phone:
- Sensor Data from Accelerometer
- Sensor Data from GPS sensors
- Daily Human Activities (walking, standing, sitting, etc.)
- Used Service Lines (bus, train, tram)

All collected data is transferred to our data center. The Live+Gov Services does NOT store or collect names and email addresses.

Which processing is applied to the data?
We process Accelerometer samples on the Mobile device in order to extract human activities. We process GPS coordinates on the Live+Gov server to extract the currently used service lines.

The stored information is used to produce aggregate reports. These reports are based on anonymized data that do not allow the reconstruction of the routes of individual persons at a given point in time.

Access to Stored Data
All personal data can be accessed at our Privacy Dashboard at http://liveandgov.uni-koblenz.de/trial/dashboard

There you have the possibility to:
- View all recordings made
- Delete individual recordings
- View the raw data collected for each recording
- Download the data for each recording
- View the results of the applied data processing

When is the data deleted?
The full data-set will be stored until January 2016.

Anonymized artifacts of this data are stored longer than that and will be used by the Consortium Partners for research purposes.

For which purposes is the data collected?
The collected data will be used for personalization of user experience of the mobile application. Moreover, we will analyze travel patterns in the Helsinki Urban region for

general research and for improvement and optimization of the infrastructure provided by HSL.

Collaboration with Third Parties
The raw data is not shared with any parties external to the Live+Gov Consortium.

Who has access to the data? The following organizations inside the Live+Gov Consortium have access to the collected data:

1. University of Koblenz-Landau
   http://www.uni-koblenz-landau.de/
   Universittsstrae 1
   56070 Koblenz

2. Mattersoft
   http://www.mattersoft.fi/
   Hmeenkatu 13, 33100 Tampere, Finland
   +358 10 3225000

3. Centre for Research and Technology Hellas (CERTH)
   http://mklab.iti.gr/contact
   6th km Charilaou-Thermi Road
   P.O. Box 60361,
   57001 Thermi-Thessaloniki, Greece

Aggregated report based on anonymized data are disclosed to officials at HSL http://www.hsl.fi/ and might be made public in the form of a blog post or research article. These aggregated reports show distributions of all routes that have been collected in the system and do not allow to infere the location of a single user at a given point in time.

Identity of Data Controller
The data is controlled by Dr. H. Hartmann <hartmann@uni-koblenz.de>.

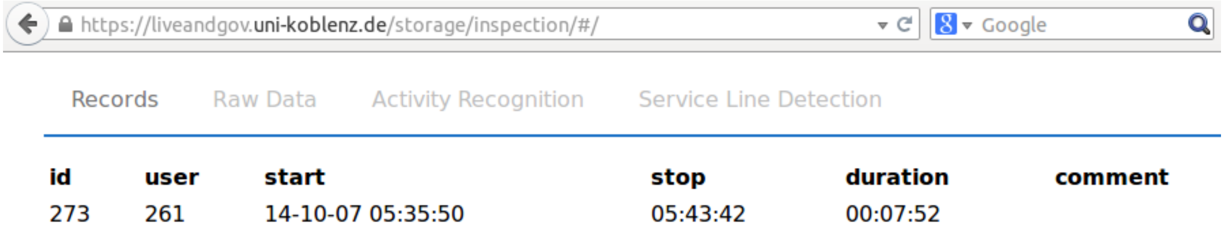
Consent
I have read and understood the above privacy policy and agree that my data is collected and processed accordingly.

4.3 Transfer Encryption (HTTPS/SSL)

The data transfer from the mobile device to the Data Storage component as well as the transfer to the Privacy Dashboard relay on HTTP. In order to avoid eavesdropping by externals that have access to the communication channels, we secured all HTTP connections with SSL encryption. This measure implements the security recommendation R7 Securing Communiction Channels of Section 3.2.2.3.

The implementation of this measure was rather straight forward. A SSL certificate was issued by the University of Koblenz, and installed to our webservers. Requests to this servers can now be done using HTTPS. Figure 15 shows a screenshot of such a secured connection to our Privacy

Dashboard. The necessary adjustments to the Sensor Collection Component were minimal, since Android comes with native support for HTTPS connections.



id	user	start	stop	duration	comment
273	261	14-10-07 05:35:50	05:43:42	00:07:52	

Figure 15: HTTPS connection to Privacy Dashboard

4.4 Security of Infrastructure

The servers hosting the Sensor Data Storage component have been secured using the following security measures:

- Firewall. A Firewall restricts access to the minimal necessary ports allowing upload and inspection of the data as well, database access by our consortium partners and maintenance over SSH.
- Security Upgrades. The operating system (Ubuntu Linux 12.04) is configured to install security updates of all installed packages on a daily basis.
- Access Rules. Access to the System (via SSH) and the database is restricted to a minimal amount of persons. Those persons have been advised to use strong passwords.

These measures are a basic implementation of recommendation R6.

4.5 Data Expiring Dates

Trips stored in the Sensor Data Storage Service have been endowed with a data expiration date. A python script is provided that performs the of expired data. In our installation we have this script running as a cron job every night. The resulting schema of the trip table listed in Figure 16. The expiring date is encoded as a 64bit integer holding the unix timestamp in seconds of the expiration date.

By default the expiration data is set to one year after the recording was uploaded. We are working on including a feature for setting the expiration date for each recording inside the privacy dashboard.

4.6 Pseudonymization of trips.

The Sensor Data Storage Service follows a privacy by design guideline in that it does not store personal information like names and email address at all, only pseudonyms in the form of user-ids.

```
  Column   |          Type
-----------+------------------------
 trip_id   | integer
 user_id   | character varying(36)
 start_ts  | bigint
 stop_ts   | bigint
 name      | character varying(255)
 deleted   | boolean
 expires   | bigint
```

Figure 16: Database Schema of the Trip Table

In the case that such information should be linked to the user-id stored in the trip table we provide a python script (cf. Figure 17) that randomizes the user-ids and thereby removes all links that might have existed before. The scripts allows as a parameter a general SQL query that selects a sub-set of the trips that shall be anonymized.

This measure implements part of the security recommendation R5 Anonymize personal data of Section 3.2.2.3.

4.7 GPS Anonymization: Gaussian Noise

The Gaussian Noise anonymization method adds random noise to the stored GPS points. The `privacy.py` script (cf. Figure 17) implements this feature as ”blur” and allows the specification of the displacement variance. Figure 18 shows an example of a GPS track with and without added noise.

The Gaussian Noise anonymization, even with a low noise radius of 5m, disables the possibility to correlate GPS data with data from cameras to identify a single person. It is typically used in conjunction with other method to provide a maximum of anonymity.

4.8 GPS Anonymization: K-Anonymity

This method of GPS anonymization is a variant of Sweeney’s K-Anonymity [17] applied to GPS data along the lines of Grutser and Grunwald [11]. The basic idea is that the resulting data from a single person should not distinguishable form data from K-1 other persons. In order to achieve such kind of anonymization we calculate for each GPS point P by a user U the $K-1$ nearest GPS points $Q_1,\dots,Q_{K-1}$ which are recorded by $K-1$ different users. The anonymized point R is now the centroid of the K points $(P,Q_1,\dots,Q_{K-1})$. This operations applied to all data points P in the dataset. The resulting anonymized points R are stored in a new table, so that these points do not take part in further the centroid calculations.

Figure 19 shows the results of application of K-anonymity to the original GPS trip displayed in Figure 18. K-Anonymity implementation script (cf. Figure 17) supports a second parameter that specifies the maximal distance that a point can get displaced by this method.

This measure implements part of the security recommendation R5 Anonymize personal data of Section 3.2.2.3.

```
Usage: privacy.py [-d database name] [-u database user]
                  [-p database password]
                  [-h database host] [-o operation]

Valid Operations:

cleanup: Flags all expired rows as deleted
    No Options

anonymization: Randomizes user_ids
    Option:
        SQL Select Query: Every row found by this query will
        be randomized.

haircut: Cuts trips where they split and could be traceable
    Options:
        First - the radius in which we look
        Second - the number of users that needs to be in the
                 radius

blur: Offsets every GPS point by a random amount
    Options:
        The expected distance a point has to its origin

k_anonymity: Creates a centroid for every GPS point
    Options:
        First - Number of maximum GPS points to use. If 0 then
                we grab every point inside the radius
        Second - Maximum distance a point can have to be used
                 for the creation of the new point
```

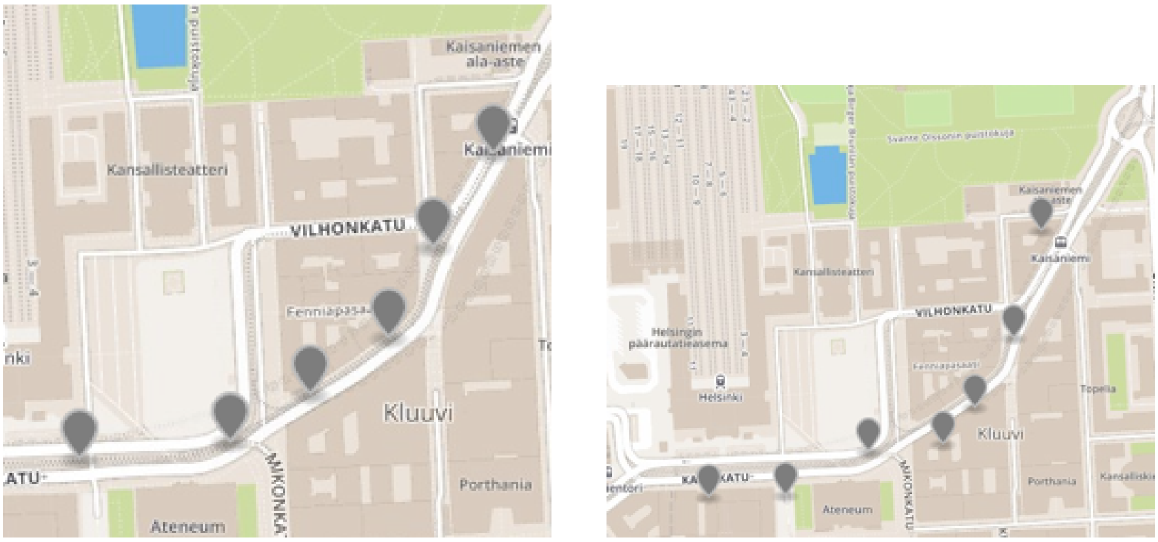
Figure 17: Usage options for the privacy.py script.

4.9 GPS Anonymization: Haircut

The *haircut* anonymization method was developed by UKob to address the problem of GPS tracks that show locations, like the home address, that are unique to a single user. Looking on our data we see many trips that originate at a unique location and later on meet other tracks at locations of public interest like bus stops or larger streets. The haircut methods cuts away parts of tracks that ”stick-out” the whole set of recorded trajectories. For a given GPS location *P* the number of users is computed that have GPS samples near to *P*. If this number is lower than a given threshold (e.g. 3) the point *P* is dropped from the dataset.

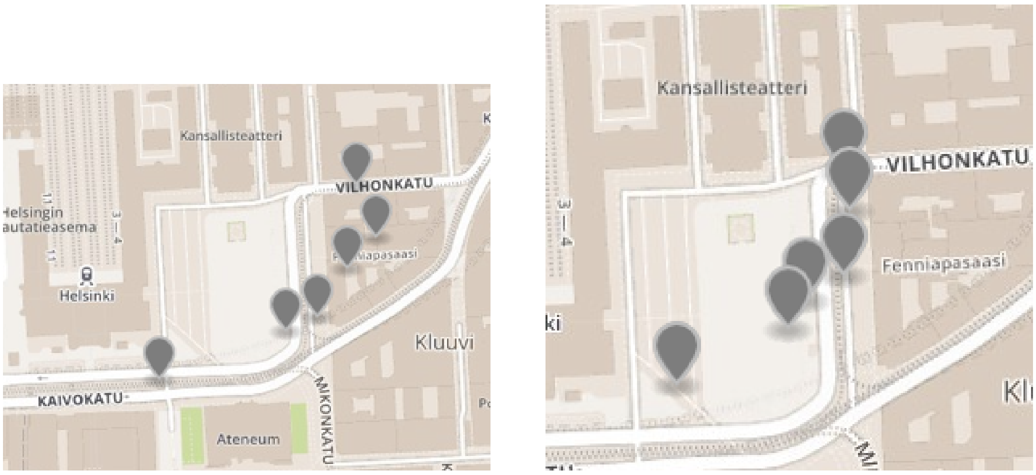
The implementation script (cf. Figure 17) takes two parameters that specify the radius to look for other GPS samples and the number of different users that are required to have samples inside this radius. Figure 20 shows an example a two-trips before and after the haircut.

This measure implements part of the security recommendation R5 Anonymize personal data of Section 3.2.2.3.



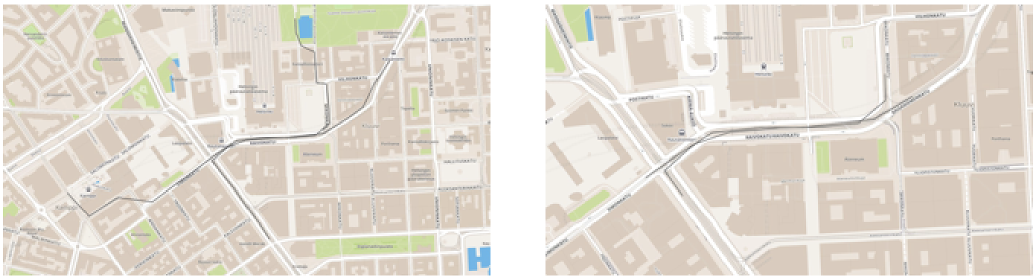
(a) Original GPS samples (b) GPS samples with noise

Figure 18: GPS Anonymization with Gaussian Noise



(a) K=7 anonymization (b) K=15 Anonymization

Figure 19: GPS Anonymization with K-Anonimity



(a) Original GPS tracks (b) GPS tracks with haircut

Figure 20: GPS Anonymization with Haircut Method

5 Conclusion

In this deliverable we have presented a thorough analysis of the legal, ethical and practical aspects of privacy surrounding the Live+Gov data mining scenarios.

The main finding is that privacy is the control that citizens have over their private data. Following this definition, there are two different ways to enhance the privacy awareness of an IT system. The first one, is to make the data non-personal by using anonymization, the other is to give the citizen the control over their personal data.

In the context of sensor data the first approach is inherently difficult, since the data variety is so high and highly unique so that an identification of the citizen that recorded the data can never be completely ruled out. The presented anonymization implementations for GPS data are best-effort methods to avoid such identifications.

The main measures to preserve the citizens privacy have thus focused on the second approach of giving citizens control over their data. In this deliverable we have presented a ”Privacy Dashboard” that gives the citizen a very detailed overview over the stored and processed data.

6 References

[1] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr. Basic concepts and taxonomy of dependable and secure computing. *Dependable and Secure Computing, IEEE Transactions on*, 1(1):11–33, Jan 2004.

[2] Peter Bodorik and Dawn N. Jutla. Sociotechnical architecture for online privacy. *Security & Privacy*, 3:29–39, 2005.

[3] Chris Chambers. Nsa and gchq: the flawed psychology of government mass surveillance, 2013. http://www.theguardian.com/science/head-quarters/2013/aug/26/nsa-gchq-psychology-government-mass-surveillance.

[4] Roger Clarke. Introduction to dataveillance and informatin privacy, and defintions of terms, 1997. http://www.rogerclarke.com/DV/Intro.html.

[5] European Commission. 2000/520/ec: Commission decision of 26 july 2000 (safe habor), 2000. http://eur-lex.europa.eu/.

[6] Judith DeCew. Privacy. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2013 edition, 2013. http://plato.stanford.edu/archives/fall2013/entries/privacy/.

[7] Rachel L. Finn, David Wright, and Michael Friedewald. Seven types of privacy. In *European Data Protection: Coming of Age*. Springer Netherlands, 2013.

[8] European Union Agency for Fundamental Rights. Handbook on european data protection law, 2014. http://fra.europa.eu/.

[9] Charles Fried. Privacy: A rational context. In *An Anatomy of Values*, pages 137–152. Havard Univ. Press, 1970.

[10] Rüdiger Grimm, Daniela Simić-Draws, Katharina Bräunlich, Andreas Kasten, and Anastasia Meletiadou. Referenzmodell für ein vorgehen bei der it-sicherheitsanalyse. *Informatik-Spektrum*, 2014.

[11] Marco Gruteser and Dirk Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services*, MobiSys '03, pages 31–42, New York, NY, USA, 2003. ACM.

[12] Serge Gutwirth. Privacy and the information age, 2002.

[13] John Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6):391–399, 2009.

[14] OECD. Recommendations of the council concerning guidelines governing the protection of privacy and trans-border flows of personal data, 1980. http://www.oecd.org/.

[15] Council of Europe. European convention for the protection of human rights and fundamental freedoms, as ammended by protocols nos. 11 and 14, 1950. http://conventions.coe.int/.

[16] Council of Europe. Convention for the protection of individuals with regard to automatic processing of personal data, 1981. http://conventions.coe.int/.

[17] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[18] European Union. 2007 pnr agreement, 2007. http://eur-lex.europa.eu/.

[19] European Union. Charter of fundamental rights of the european union, 2010. http://eur-lex.europa.eu/.

[20] European Unioon. Directive 95/46/ec of the european parliament and of the council of the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995. http://eur-lex.europa.eu/.